

Dark census: Statistically detecting the satellite populations of distant galaxies

Francis-Yan Cyr-Racine,^{1,2,3} Leonidas A. Moustakas,^{2,3} Charles R. Keeton,⁴ Kris Sigurdson,⁵ and Daniel A. Gilman⁶¹*Department of Physics, Harvard University, Cambridge, Massachusetts 02138, USA*²*NASA Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91109, USA*³*California Institute of Technology, Pasadena, California 91125, USA*⁴*Department of Physics and Astronomy, Rutgers, The State University of New Jersey, Piscataway, New Jersey 08854, USA*⁵*Department of Physics and Astronomy, University of British Columbia, Vancouver, British Columbia V6T 1Z1, Canada*⁶*Department of Physics and Astronomy, University of California, Los Angeles, California 90095, USA*

(Received 18 June 2015; published 5 August 2016)

In the standard structure formation scenario based on the cold dark matter paradigm, galactic halos are predicted to contain a large population of dark matter subhalos. While the most massive members of the subhalo population can appear as luminous satellites and be detected in optical surveys, establishing the existence of the low mass and mostly dark subhalos has proven to be a daunting task. Galaxy-scale strong gravitational lenses have been successfully used to study mass substructures lying close to lensed images of bright background sources. However, in typical galaxy-scale lenses, the strong lensing region only covers a small projected area of the lens's dark matter halo, implying that the vast majority of subhalos cannot be directly detected in lensing observations. In this paper, we point out that this large population of dark satellites can collectively affect gravitational lensing observables, hence possibly allowing their statistical detection. Focusing on the region of the galactic halo outside the strong lensing area, we compute from first principles the statistical properties of perturbations to the gravitational time delay and position of lensed images in the presence of a mass substructure population. We find that in the standard cosmological scenario, the statistics of these lensing observables are well approximated by Gaussian distributions. The formalism developed as part of this calculation is very general and can be applied to any halo geometry and choice of subhalo mass function. Our results significantly reduce the computational cost of including a large substructure population in lens models and enable the use of Bayesian inference techniques to detect and characterize the distributed satellite population of distant lens galaxies.

DOI: [10.1103/PhysRevD.94.043505](https://doi.org/10.1103/PhysRevD.94.043505)

I. INTRODUCTION

Dark matter forms the gravitational backbone of most of the observed structures in the Universe, from the largest galaxy clusters to the faintest dwarf galaxies. Despite this ubiquity, the nature of dark matter remains a mystery. On the one hand, the cold dark matter (CDM) paradigm has been extremely successful at describing observations on large cosmological scales such as the cosmic microwave background [1], the clustering of galaxies [2], and cosmic shear measurements [3]. On the other hand, this success constitutes a mixed blessing since there is a vast array of particle candidates that naturally fall under the CDM umbrella on large cosmological scales. One possible avenue to distinguish between this plethora of models is to look on much smaller length scales where clues of the particle nature of dark matter are more evident. For instance, the initial free-streaming of warm dark matter particle would damp the growth of structure on small scales [4–11], while the fluctuation power spectrum of dark matter that couples to relativistic species until late times would display both acoustic oscillations and collisional damping

[12–17]. On the other hand, if these physical phenomena are absent in the dark matter sector, standard structure formation theory predicts that galaxies should harbor a very large number of dark satellites [18,19]. The statistical detection of these numerous dark subhalos would validate a key prediction of standard CDM theory.

Strong gravitational lensing provides a way to probe the distribution of dark matter on the smallest scales [20–86]. For instance, the observations of flux-ratio anomalies in multiply imaged lensed quasars have been used to study of abundance of mass substructures within the lens galaxy [20,34,74,79,80,82,84–86]. More recently, direct gravitational imaging [21,28,36,43,87] has enabled the detection of massive substructures along magnified arcs and Einstein rings. Similarly, resolved spectroscopy of strongly lensed dusty star-forming galaxies has been proposed to study mass substructures within the lens galaxy [24,88,89]. By construction, these techniques are most sensitive to substructures lying close to lensed images, that is, substructures that appear close in projection to the Einstein radius of the lens. Since the typical Einstein radius of a galaxy-scale

lens is a small fraction of its virial radius, only a small number of mass substructures are on average projected close to the region probed by strong lensing [31]. One therefore naturally expects that an order unity number of mass substructures could be detectable in each individual lens. Meaningful constraints on mass substructure inside lens galaxies can then be obtained by considering a sample of galaxies as was recently done in Ref. [21].

While the vast majority of mass substructures in lens galaxies cannot be directly detected in lensing observations, the collective effect of substructures far from lensed images can nevertheless cause small perturbations to lensing observables that can be statistically detected. For instance, Ref. [50] studied how astrometric perturbations could be used to probe mass substructures, while Ref. [30] used both astrometric and magnification perturbations to constrain the presence of dark clumps within the lens HE0435-1223. In addition, time-delay fluctuations in multiply imaged lensed quasars have been proposed [42] as a tool to characterize broad properties of mass substructures within lens galaxies. Certainly, the overall population of mass substructures will perturb all lensing observables in a coherent and correlated way.

In this manuscript, we develop a formalism to study stochastic millilensing from a large population of unresolved mass substructures inside the halos of galaxies acting as strong gravitational lenses. The aim of this formalism is to compute the joint effect of mass substructures on all lensing observables (image positions, magnifications, time delays), taking into account all possible correlations among those. Our work builds on the theory of stochastic stellar microlensing [90–100] and generalizes the results of Refs. [38,40,41]. We focus our analysis on multiply imaged point sources (e.g. quasar lenses) since these are the most relevant objects where gravitational time delays can in principle be measured. As we discuss below, time-delay measurements are crucial in probing the satellites populating the outskirts of distant lens galaxies.

As in some of the stellar microlensing works, we use Markov’s method (see, e.g., Ref. [101]) to compute from first principles the probability distribution of lensing potential and deflection perturbations in the presence of a population of mass substructures inside the lens galaxy. By performing an Edgeworth expansion [102], we show that for a realistic structure formation scenario the probability distributions are nearly Gaussian. We also compute the leading order deviations from pure Gaussianity. The advantage of our analysis is that it allows one to determine which physical quantities control the behavior of lensing observables in the presence of mass substructures. This dependence on physical parameters is often obscured in studies relying purely on numerical simulations. By removing the need to perform such simulations, our approach has the potential to significantly speed up the analysis of substructures inside lens galaxies and provide a convenient

way to explore degeneracies between the macrolens and the substructure population. Most importantly, it provides a practical way to statistically detect dark satellites inside lens galaxies, hence providing a key test of standard cold dark matter theory.

In this paper, we focus on analyzing the effect of mass substructures that are spatially located beyond a few Einstein radii. We leave the analysis of local mass substructures that are spatially located close to lensed images to future work. This paper is organized as follows. In Sec. II, we describe the challenge of statistically detecting a population of unresolved mass substructures and explain our approach to tackle this problem via the use of the characteristic function. In Sec. III, we justify the division of the overall substructure population into two subpopulations (distributed and local), and we perform the actual calculation of the characteristic function for a population of distributed substructures. We show that in the cases of interest its behavior is quasi-Gaussian, and we discuss in which situations non-Gaussianities can become important. We also compare our results to the output of Monte Carlo realizations. We then show in Sec. IV how our approach can be used to marginalize over the distributed substructure population. We discuss which physical properties of the distributed substructure population are most relevant to the lensing observables in Sec. V, and we finally conclude in Sec. VI.

II. STOCHASTIC LENSING: GENERAL CASE

In this section, we present the general ideas behind our approach to substructure lensing. After brief remarks about our setup and notation, we introduce the challenges of lens modeling in the presence of a stochastic population of mass substructures. We then present the basic ideas behind our analytical stochastic lensing framework and derive important results regarding the joint distribution of gravitational lensing observables. These results are used throughout the rest of this paper.

A. Setup and notation

For definiteness, we consider a situation where a high-redshift point-like source (such as a quasar) is multiply imaged by a massive foreground galaxy whose properties are described by a set of parameters \mathbf{q}_{gal} . For instance, \mathbf{q}_{gal} could contain information about the lens Einstein radius, the dark matter and stellar density profiles, their ellipticity, etc. The characteristics of the main lens can also depend on the fundamental properties of dark matter (denoted by the parameters \mathbf{q}_{DM}) such as its free-streaming length (λ_{fs}), its sound horizon (r_{DAO}), its self-interaction cross section (σ_{SIDM}), and its temperature of kinematic decoupling (T_{kd}). In addition, the main lens galaxy lives in a local environment characterized by parameters \mathbf{q}_{env} which contain information, for instance, about the external shear and the properties of nearby luminous galaxies. We parametrize

TABLE I. Summary of our notation for the different sets of parameters relevant to our gravitational lensing analysis. The third column indicates the relative dependency of the different sets of parameters, while the last column gives examples of the different types of parameters. See main text for more details.

Parameters	Description	Dependency	Example
\mathbf{q}_{cos}	Cosmological parameters	\dots	$H_0, \Omega_m, \Omega_\Lambda, A_s, n_s$
\mathbf{q}_{DM}	Dark matter parameters	\mathbf{q}_{cos}	$\lambda_{\text{fs}}, r_{\text{DAO}}, \sigma_{\text{SIDM}}, T_{\text{kd}}$
\mathbf{q}_{env}	Lens environment parameters	\mathbf{q}_{cos}	External shear, positions, and mass of nearby galaxies
\mathbf{q}_{gal}	Macrolens parameters	$\mathbf{q}_{\text{cos}}, \mathbf{q}_{\text{DM}}, \mathbf{q}_{\text{env}}$	Lens Einstein radius, ellipticity, density profile
\mathbf{q}_{los}	Line-of-sight structure parameters	$\mathbf{q}_{\text{cos}}, \mathbf{q}_{\text{DM}}$	Nonlinear matter power spectrum parameters
\mathbf{q}_{sub}	Substructure population parameters	$\mathbf{q}_{\text{cos}}, \mathbf{q}_{\text{DM}}, \mathbf{q}_{\text{gal}}$	Substructure mass function and their spatial distribution
\mathbf{c}_{sub}	Individual substructure parameters	\mathbf{q}_{sub}	Positions, masses, and concentrations of each subhalo

the line-of-sight structures (that is, exterior to the main lens plane) between the source and the observer via an array \mathbf{q}_{los} . Of course, all of these different sets of parameters have a dependence on the background cosmology, which we denote as $\mathbf{q}_{\text{cos}} = \{H_0, \Omega_m, \Omega_\Lambda, A_s, n_s\}$, where H_0 is the Hubble parameter, Ω_m and Ω_Λ are the energy densities in matter and dark energy, respectively, in units of the critical energy density of the Universe, A_s is the amplitude of the primordial power spectrum of scalar fluctuations, and n_s is the scalar spectral index. Throughout this paper, we use the numerical values for the cosmological parameters from the Planck 2015 data release [103].

In addition to the spatially smooth component described by \mathbf{q}_{gal} , the lens galaxy also contains mass substructures, the most luminous of which can appear as satellites orbiting the main lens. We collect the individual properties of these substructures in an array \mathbf{c}_{sub} , which could, for instance, contain information about the position, virial mass, and concentration of each substructure. Finally, we assume that individual substructure properties are sampled from an underlying probability distribution parametrized by an array \mathbf{q}_{sub} which encodes information about the substructure mass function, their spatial distribution within the lens, and their mass-concentration relation, which has a strong dependence on the parameters contained in \mathbf{q}_{DM} . We summarize our notation in Table I and indicate the interdependency of these different sets of parameters. Also, we refer the reader to Tables II, III, and IV of Appendix E for details about the notation used throughout this paper.

We reiterate that our goal is to use gravitational lenses to constrain the substructure population parameters \mathbf{q}_{sub} and then use this information to learn new insights about the microphysics of dark matter encoded in \mathbf{q}_{DM} . Of course, determining the impact of a given choice of \mathbf{q}_{cos} , \mathbf{q}_{DM} , and \mathbf{q}_{gal} on the population parameters \mathbf{q}_{sub} is highly nontrivial and requires detailed numerical simulations. This is a very active area of research, and tremendous progress has been made in the last decade [104–120]. In this work, we are interested in developing a formal understanding of the impact of a given choice of \mathbf{q}_{sub} on lensing observations. We defer to future work the interpretation of given \mathbf{q}_{sub} constraints in terms of dark matter microphysics.

B. Stochastic lensing by unresolved substructures: Main challenge

In this subsection, we review the challenges of lens modeling in the presence of unresolved mass substructures. Let us imagine that we have a data vector \mathbf{d} . In general, \mathbf{d} could include the position and surface brightness of the multiple images of the source, the time delays between the images, the velocity dispersion of the lens, and other data about the lens environment (external shear and convergence). Using these data, we would like to jointly constrain the arrays of parameters $\mathbf{q} \equiv \{\mathbf{q}_{\text{gal}}, \mathbf{q}_{\text{env}}, \mathbf{q}_{\text{los}}\}$ and \mathbf{c}_{sub} . Given a choice of these parameters, one can compute the theory observables $\mathbf{t}(\mathbf{q}, \mathbf{c}_{\text{sub}})$ (encompassing, for instance, image magnifications and positions, as well as relative time delay between lensed images) and use them to compute the likelihood of measuring \mathbf{d} , $\mathcal{L}(\mathbf{d}|\mathbf{q}, \mathbf{c}_{\text{sub}})$. The problem with the above procedure is that a given dark matter theory does not predict the positions and masses of individual subhalos within the lens galaxy. The fundamental dark matter physics rather makes predictions about the statistical properties of subhalos (described here by \mathbf{q}_{sub}) such as their mass distribution, their concentration, and their spatial distribution within the lens. Therefore, the elements of array \mathbf{c}_{sub} are nuisance parameters that need to be integrated out.

One could however sidestep this issue by directly specifying the *statistics* of the substructure population via the set of parameters \mathbf{q}_{sub} , without having to explicitly draw random realizations \mathbf{c}_{sub} . The immediate problem with this approach is that the theory observables are no longer unambiguously specified. Instead, for each choice of substructure population parameters \mathbf{q}_{sub} , one obtains a probability density function for the theory observables $P(\mathbf{t}|\mathbf{q}, \mathbf{q}_{\text{sub}})$. Formally, this probability density can be written as

$$P(\mathbf{t}|\mathbf{q}, \mathbf{q}_{\text{sub}}) = \int P_{\text{sub}}(\mathbf{c}_{\text{sub}}|\mathbf{q}_{\text{sub}}) \delta_D^k(\mathbf{t} - \mathbf{t}(\mathbf{q}, \mathbf{c}_{\text{sub}})) d\mathbf{c}_{\text{sub}}, \quad (1)$$

where $P_{\text{sub}}(\mathbf{c}_{\text{sub}}|\mathbf{q}_{\text{sub}})$ is the probability of having a mass substructure population specified by \mathbf{c}_{sub} , given a choice of population parameters \mathbf{q}_{sub} , and where δ_D^k is the

k -dimensional Dirac delta function (k is the length of the \mathbf{t} vector). Once $P(\mathbf{t}|\mathbf{q}, \mathbf{q}_{\text{sub}})$ is specified, the likelihood of the data \mathbf{d} now takes the form,

$$\mathcal{L}(\mathbf{d}|\mathbf{q}, \mathbf{q}_{\text{sub}}) = \int d\mathbf{t} P(\mathbf{t}|\mathbf{q}, \mathbf{q}_{\text{sub}}) \mathcal{L}(\mathbf{d}|\mathbf{t}), \quad (2)$$

where $\mathcal{L}(\mathbf{d}|\mathbf{t})$ is the likelihood of the data given the theory observables. Note that if we substitute Eq. (1) into Eq. (2), we obtain

$$\mathcal{L}(\mathbf{d}|\mathbf{q}, \mathbf{q}_{\text{sub}}) = \int P_{\text{sub}}(\mathbf{c}_{\text{sub}}|\mathbf{q}_{\text{sub}}) \mathcal{L}(\mathbf{d}|\mathbf{q}, \mathbf{c}_{\text{sub}}) d\mathbf{c}_{\text{sub}}, \quad (3)$$

which is just the standard marginalization over the substructure population. Once $\mathcal{L}(\mathbf{d}|\mathbf{q}, \mathbf{q}_{\text{sub}})$ is known, it is straightforward to construct the desired posterior distribution $P(\mathbf{q}, \mathbf{q}_{\text{sub}}|\mathbf{d}) \propto \mathcal{L}(\mathbf{d}|\mathbf{q}, \mathbf{q}_{\text{sub}}) \Pi(\mathbf{q}, \mathbf{q}_{\text{sub}})$, where $\Pi(\mathbf{q}, \mathbf{q}_{\text{sub}})$ is the prior probability distribution for \mathbf{q} and \mathbf{q}_{sub} .

The above calculation of $P(\mathbf{q}, \mathbf{q}_{\text{sub}}|\mathbf{d})$ hinges on the accurate determination of the likelihood $\mathcal{L}(\mathbf{d}|\mathbf{q}, \mathbf{q}_{\text{sub}})$, either through Eq. (2), or directly through Eq. (3). Let us for now focus on the latter approach which has been used in various works on mass substructure inside gravitational lenses [20,34,74,79,82,84–86]. If one could rapidly draw a large number of substructure realizations \mathbf{c}_{sub} from the distribution $P_{\text{sub}}(\mathbf{c}_{\text{sub}}|\mathbf{q}_{\text{sub}})$ and compute the theory observables $\mathbf{t}(\mathbf{q}, \mathbf{c}_{\text{sub}})$ for each of those, one could then replace the integral in Eq. (3) by a sum of all the realizations

$$\mathcal{L}(\mathbf{d}|\mathbf{q}, \mathbf{q}_{\text{sub}}) \approx \sum_{\mathbf{c}_{\text{sub}} \sim P_{\text{sub}}(\mathbf{q}_{\text{sub}})} \mathcal{L}(\mathbf{d}, \mathbf{t}(\mathbf{q}, \mathbf{c}_{\text{sub}})), \quad (4)$$

where the notation $\mathbf{c}_{\text{sub}} \sim P_{\text{sub}}(\mathbf{q}_{\text{sub}})$ means that \mathbf{c}_{sub} is drawn from the distribution $P_{\text{sub}}(\mathbf{q}_{\text{sub}})$. This approach has several drawbacks. First, it is difficult to assess how many realizations are needed to properly estimate the likelihood. A related issue is how to identify the substructure realizations that contribute most to the sum and make sure that these realizations are included in it. Second, for dark matter models that predict an abundance of subhalos, randomly drawing a realization can be a numerically costly process since thousands or millions of subhalos need to be included in the lensing calculation. Most importantly, a purely numerical approach obscures which key physical quantities control the impact of substructures on lensing observables. While this approach is viable if we are interested in accurately knowing the likelihood for a few points in parameter space, it is impractical if we are using a Markov Chain Monte Carlo approach to estimate the final posterior distribution of \mathbf{q} and \mathbf{q}_{sub} . To remedy these problems, we describe in the following section an approach that allows efficient computation of stochastic lensing observables while leaving the physics of substructure lensing transparent.

C. Stochastic substructure lensing: Characteristic function approach

We now turn our attention to an analytical approach to the computation of lensing observables in the presence of a population of unresolved mass substructures. The calculation presented here draws from the theory of stochastic microlensing in the presence of a uniform star field [90–100]. As a starting point, our technique takes full advantage of two simplifying facts about the impact of mass substructures on the lensing observables:

- (i) For deflection, shear, convergence, and projected gravitational potential, the overall impact of the subhalo population is the sum of the contributions from each mass substructure.
- (ii) Each subhalo is independent of all other subhalos in the lens.

The first assumption is always true and is a direct consequence of the linearity of Poisson's equation. The second assumption is not strictly true since mass substructures may be themselves clustered within galactic halos. However, the relative importance of substructure clustering will be diminished by projection effects since lensing is only sensitive to the mass distribution integrated along the line of sight. Moreover, tidal interactions between subhalos and the tidal field of the host galaxy will tend to erase correlations among substructures within a few dynamical times [121]. Thus, to a good approximation, we can use the above simplifying facts to make progress in evaluating $P(\mathbf{t}|\mathbf{q}, \mathbf{q}_{\text{sub}})$. For the moment, let us focus on the lensing quantities that receive purely additive corrections from the substructures. These include the projected gravitational potential ϕ , the deflections $\vec{\alpha} = \vec{\nabla}\phi$, the convergence $\kappa = (\phi_{xx} + \phi_{yy})/2$, as well as the shears $\gamma_c = (\phi_{xx} - \phi_{yy})/2$ and $\gamma_s = \phi_{xy}$. We denote these “linear” lensing quantities¹ by $\mathbf{t}_L = \{\dots, \{\phi^{(j)}, \vec{\alpha}^{(j)}, \kappa^{(j)}, \gamma_c^{(j)}, \gamma_s^{(j)}\}, \dots\}$, where the index j denotes that these quantities are evaluated at the position of the j th image. Our goal is to take advantage of the linearity to first compute $P(\mathbf{t}_L|\mathbf{q}, \mathbf{q}_{\text{sub}})$ and then reconstruct $P(\mathbf{t}|\mathbf{q}, \mathbf{q}_{\text{sub}})$ using the relation

$$P(\mathbf{t}|\mathbf{q}, \mathbf{q}_{\text{sub}}) = \int d\mathbf{t}_L P(\mathbf{t}_L|\mathbf{q}, \mathbf{q}_{\text{sub}}) \delta_D(\mathbf{t} - \mathbf{t}(\mathbf{t}_L)). \quad (5)$$

The linear quantities \mathbf{t}_L receives contribution from both the smooth mass model and its environment (described by \mathbf{q}) and the mass substructures themselves

$$\mathbf{t}_L(\mathbf{q}, \mathbf{c}_{\text{sub}}) = \bar{\mathbf{t}}_L(\mathbf{q}) + \sum_{i=1}^N \delta \mathbf{t}_L^{(i)}, \quad (6)$$

¹Note that we use the nomenclature “quantities” and not “observables” since ϕ , $\vec{\alpha}$, κ , γ_c , and γ_s are not directly observable.

where $\bar{\mathbf{t}}_L(\mathbf{q})$ describes the contribution from the smooth model and its environment, while $\delta\mathbf{t}_L^{(i)} \equiv \delta\mathbf{t}_L(\mathbf{q}, \mathbf{c}_i)$ is the contribution from mass substructure i . Here, N is the total number of subhalos within the lensing galaxy. Since $\bar{\mathbf{t}}_L(\mathbf{q})$ is completely fixed by a given choice of \mathbf{q} , the inherent stochasticity of \mathbf{t}_L is entirely caused by the random mass substructures inside the lens galaxy. Defining $\Delta\mathbf{t}_L \equiv \sum_{i=1}^N \delta\mathbf{t}_L^{(i)}$, all relevant statistical information about the mass substructures is contained in the probability density function $\Phi_N(\Delta\mathbf{t}_L|\mathbf{q}, \mathbf{q}_{\text{sub}})$ for the collective effect $\Delta\mathbf{t}_L$ of N substructures on the linear lensing quantities. Once Φ_N is known, the probability density $P(\mathbf{t}_L|\mathbf{q}, \mathbf{q}_{\text{sub}})$ appearing in Eq. (5) is simply given by

$$P(\mathbf{t}_L|\mathbf{q}, \mathbf{q}_{\text{sub}}) = \Phi_N(\mathbf{t}_L - \bar{\mathbf{t}}_L(\mathbf{q})|\mathbf{q}, \mathbf{q}_{\text{sub}}). \quad (7)$$

Effectively, $\Phi_N(\Delta\mathbf{t}_L|\mathbf{q}, \mathbf{q}_{\text{sub}})$ is an l -dimensional joint probability distribution for l sums of N independent random variables, where l refers to the number of linear lensing quantities included in the analysis. Take $\Phi_1(\delta\mathbf{t}_L^{(i)}|\mathbf{q}, \mathbf{q}_{\text{sub}})$ to be the joint probability distribution for the linear lensing quantities in the presence of a *single* substructure. For now, we assume that we know the functional form of $\Phi_1(\delta\mathbf{t}_L^{(i)}|\mathbf{q}, \mathbf{q}_{\text{sub}})$; its formal derivation is given in the next subsection. Since the subhalos are assumed to be independent of each other, $\Phi_N(\Delta\mathbf{t}_L|\mathbf{q}, \mathbf{q}_{\text{sub}})$ is given by the N -fold convolution of $\Phi_1(\delta\mathbf{t}_L^{(i)}|\mathbf{q}, \mathbf{q}_{\text{sub}})$ [101]. We then take advantage of the convolution theorem to write the characteristic function² of $\Phi_N(\Delta\mathbf{t}_L|\mathbf{q}, \mathbf{q}_{\text{sub}})$ in terms of that of $\Phi_1(\delta\mathbf{t}_L^{(i)}|\mathbf{q}, \mathbf{q}_{\text{sub}})$,

$$Q_N(\mathbf{k}_L|\mathbf{q}, \mathbf{q}_{\text{sub}}) = q_1(\mathbf{k}_L|\mathbf{q}, \mathbf{q}_{\text{sub}})^N, \quad (8)$$

where \mathbf{k}_L is the Fourier conjugate variable to $\Delta\mathbf{t}_L$, $Q_N(\mathbf{k}_L|\mathbf{q}, \mathbf{q}_{\text{sub}})$ is the characteristic function of $\Phi_N(\Delta\mathbf{t}_L|\mathbf{q}, \mathbf{q}_{\text{sub}})$, and $q_1(\mathbf{k}_L|\mathbf{q}, \mathbf{q}_{\text{sub}})$ is the characteristic function of $\Phi_1(\delta\mathbf{t}_L^{(i)}|\mathbf{q}, \mathbf{q}_{\text{sub}})$.

Now, in a typical galactic dark matter halo, the number of mass substructure N is large but unknown. Given a total convergence in dark matter substructures and a subhalo mass function, we can compute the average expected total number of mass substructures $\langle N \rangle$ [see, e.g., Eq. (21) below]. Since the evolution of mass substructures within lens galaxies is determined by the complex interplay of accretion, dynamical friction, tidal stripping, baryonic feedback, and mergers, the actual number of subhalos will typically differ from this average value. Detailed N -body simulations [122] of massive galaxies show that the scatter about the mean is consistent with that of a Poisson distribution. Then, the resulting characteristic function

²In this work, the characteristic function is simply the Fourier transform of the probability density function.

for the whole substructure population is a sum over all possible values of N , weighted by their Poisson probability with mean $\langle N \rangle$,

$$\begin{aligned} Q_{\langle N \rangle}(\mathbf{k}_L|\mathbf{q}, \mathbf{q}_{\text{sub}}) &= e^{-\langle N \rangle} \sum_{N=0}^{\infty} \frac{\langle N \rangle^N}{N!} q_1(\mathbf{k}_L|\mathbf{q}, \mathbf{q}_{\text{sub}})^N \\ &= \exp[\langle N \rangle (q_1(\mathbf{k}_L|\mathbf{q}, \mathbf{q}_{\text{sub}}) - 1)]. \end{aligned} \quad (9)$$

This result states that if one could compute $q_1(\mathbf{k}_L|\mathbf{q}, \mathbf{q}_{\text{sub}})$ for a *single* mass substructure, then one could obtain the characteristic function for the *whole population* of unresolved subhalos by taking the exponential of $\langle N \rangle (q_1(\mathbf{k}_L|\mathbf{q}, \mathbf{q}_{\text{sub}}) - 1)$. Finally, $\Phi_{\langle N \rangle}(\Delta\mathbf{t}_L|\mathbf{q}, \mathbf{q}_{\text{sub}})$ can be obtained by Fourier transforming $Q_{\langle N \rangle}(\mathbf{k}_L|\mathbf{q}, \mathbf{q}_{\text{sub}})$. Therefore, we have reduced the computation of $P(\mathbf{t}_L|\mathbf{q}, \mathbf{q}_{\text{sub}})$ for $\langle N \rangle$ subhalos to that of computing $q_1(\mathbf{k}_L|\mathbf{q}, \mathbf{q}_{\text{sub}})$ for a single substructure which is a considerable simplification.

D. Characteristic function for a single substructure

To complete our formalism, we need an expression for $q_1(\mathbf{k}_L|\mathbf{q}, \mathbf{q}_{\text{sub}})$, the characteristic function of the linear lensing quantities in the presence of a single mass substructure. We begin by writing down an expression for $\Phi_1(\delta\mathbf{t}_L|\mathbf{q}, \mathbf{q}_{\text{sub}})$,

$$\begin{aligned} \Phi_1(\delta\mathbf{t}_L|\mathbf{q}, \mathbf{q}_{\text{sub}}) &= \int P_{\text{sub}}(\mathbf{c}_{\text{sub}}^{(1)}|\mathbf{q}_{\text{sub}}) \delta_D^l(\delta\mathbf{t}_L - \delta\mathbf{t}_L(\mathbf{q}, \mathbf{c}_{\text{sub}}^{(1)})) d\mathbf{c}_{\text{sub}}^{(1)}, \end{aligned} \quad (10)$$

where $\mathbf{c}_{\text{sub}}^{(1)}$ are the parameters describing the properties of a single mass substructure. Here, $P_{\text{sub}}(\mathbf{c}_{\text{sub}}^{(1)}|\mathbf{q}_{\text{sub}})$ is the probability density function describing the probability of finding a clump of dark matter with parameters $\mathbf{c}_{\text{sub}}^{(1)}$, given a set of substructure population parameters \mathbf{q}_{sub} . The characteristic function of the above distribution, $q_1(\mathbf{k}_L|\mathbf{q}, \mathbf{q}_{\text{sub}})$, is simply the Fourier transform of Eq. (10),

$$q_1(\mathbf{k}_L|\mathbf{q}, \mathbf{q}_{\text{sub}}) = \int d\mathbf{c}_{\text{sub}}^{(1)} e^{i\delta\mathbf{t}_L(\mathbf{q}, \mathbf{c}_{\text{sub}}^{(1)}) \cdot \mathbf{k}_L} P_{\text{sub}}(\mathbf{c}_{\text{sub}}^{(1)}|\mathbf{q}_{\text{sub}}). \quad (11)$$

Computing this integral requires us to specify the spatial geometry over which the mass substructure is distributed as well as the subhalo mass function inside the lens galaxy. In the next section, we describe our strategy to evaluate this characteristic function.

III. CHARACTERISTIC FUNCTION FOR SUBSTRUCTURE POPULATION

Up to this point, we emphasize that our analysis has been very general and is purely based on the linearity and

independence of mass substructures inside galactic halos. In this section, we consider how the geometry of the substructure distribution inside galactic halos can be used to simplify the calculation of $q_1(\mathbf{k}_L|\mathbf{q}, \mathbf{q}_{\text{sub}})$. As we describe below, it is advantageous to divide the substructure population into a distributed subpopulation that contains the vast majority of subhalos and contributes small perturbations to lensing observables and into a local subpopulation that contains a few strong perturbers to lensing observables.

A. Local versus distributed substructure populations

We wish to compute the characteristic function for the linear lensing quantities in the presence of a single substructure at typical lensed image locations $\{\mathbf{x}_i\}$ situated close to the Einstein radius R_{ein} of the lens. Similar to the analysis of Ref. [38], our strategy is to divide the image plane into two regions: (i) an inner disk of radius R_{min} containing all the lensed images and a relatively small number of substructures (denoted “local” substructures) and (ii) an annulus with inner radius R_{min} and outer radius R_{max} containing the vast majority of the substructure population, which we shall refer to as the “distributed” population. This choice is illustrated in Fig. 1. The first thing that is evident from Fig. 1 is that the strong lensing region (red innermost circle) of typical galaxy-scale lenses probes the very inner part of the galactic halo. This is the region where flux ratio anomalies have been used to probe mass substructures within lens galaxies [20,34,74,79,82,84–86]. This is also the region where direct gravitational imaging [21,28,36,43,87] and spatially resolved spectroscopy [24,72,88] can be used to detect individual mass substructures within galaxy-scale lenses. In this area of the lens plane, it is possible for a mass substructure to cause a large perturbation to lensing observables which are known to cause non-Gaussian “heavy tails” [38] in stochastic lensing probability density functions. Furthermore, subhalos can have significant overlap with lensed images, implying that the internal properties of mass substructures such as their concentrations and tidal radii can be probed in this regime [20]. Due to its small size compared to the overall spatial extent of the dark matter halo, the inner region contains a relatively small fraction of the total number of mass substructures in the gravitational lens.

On the other hand, the outer region of the lens halo (the area outside the dashed blue circle) contains the vast majority of the lens galaxy mass substructures. Since they are quite distant from any lensed image, these subhalos cannot significantly affect the lensing observables on an individual basis. However, their collective effect is not necessarily negligible. Furthermore, because of their relative position with respect to the strong lensing region, we do not expect their internal structure to play a significant role in their lensing signatures. Importantly,

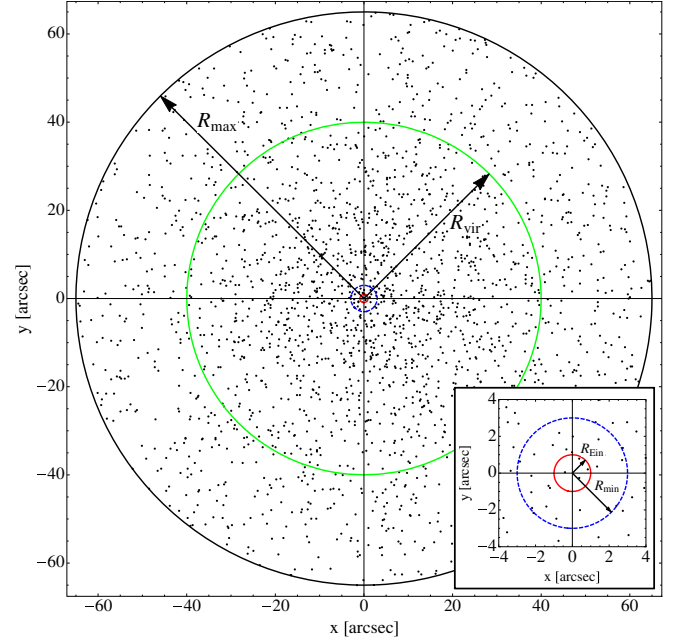


FIG. 1. Illustration of the various scales involved in galaxy-scale substructure lensing. The typical Einstein radius R_{ein} of lens galaxy (~ 1 arcsec) is indicated in red, while the typical virial radius of the galactic halo is indicated in green. The dashed blue circle and the outer black circle indicate our choice of $R_{\text{min}} (= 3R_{\text{ein}})$ and $R_{\text{max}} (= 65R_{\text{ein}})$ for the computation of the substructure characteristic function, respectively. The scattered dots represent a realization of a substructure population with the spatial distribution given in Eq. (19) with a core radius given by $r_c = 30R_{\text{ein}}$. Here, we have assumed a power law mass function as given in Eq. (20) with $\beta = -1.9$, $M_{\text{low}} = 10^7 M_{\odot}$, and $M_{\text{high}} = 10^{10} M_{\odot}$. The average convergence in mass substructure is taken to be $\langle \kappa_{\text{sub}}(R_{\text{ein}}) \rangle = 6 \times 10^{-4}$. See main text for a description of the notation. The inset at the bottom right shows an enlargement of the halo’s central region. In general, only an order unity number of substructures are projected close to the Einstein radius of the lens.

these properties of the distributed mass substructures considerably simplify the calculation of the characteristic function $q_1(\mathbf{k}_L|\mathbf{q}, \mathbf{q}_{\text{sub}})$.

It is instructive to compare the relative contribution of the distributed and local subhalo populations to the linear lensing quantities. Writing the total contribution from substructures as

$$\Delta \mathbf{t}_L = \Delta \mathbf{t}_L^{\text{local}} + \Delta \mathbf{t}_L^{\text{dist}}, \quad (12)$$

let us compare the local and distributed pieces for potential fluctuations, deflections, convergence, and shears. To do so, we generate 10^4 Monte Carlo realizations of mass substructure population. We assume the substructures to have smoothly truncated Navarro-Frenk-White (NFW) three-dimensional density profiles given by [123]

$$\rho(r_{\text{sub}}) = \frac{M_{\text{NFW}}}{4\pi r(r_{\text{sub}} + r_s)^2} \left(\frac{r_t^2}{r_{\text{sub}}^2 + r_t^2} \right), \quad (13)$$

where r_{sub} is the three-dimensional distance from the center of the subhalo, r_s is the scale radius, and r_t is the tidal radius. We note that our choice of NFW profile is more conservative than the often used Pseudo-Jaffe profile since the latter has a steeper inner density slope and has thus a larger lensing efficiency. It is important to keep in mind that observations of low-mass galaxies show mild preference for even shallower density profiles, implying that the magnitude of the local substructure perturbations discussed in this section are likely to be conservative upper bounds. For the truncated NFW profile, the mass scale M_{NFW} is related to the total mass M_{sub} of a substructure via the relation [123]

$$M_{\text{sub}} = M_{\text{NFW}} \frac{\tau^2}{(\tau^2 + 1)^2} [(\tau^2 - 1) \ln \tau + \pi \tau - (\tau^2 + 1)], \quad (14)$$

where $\tau \equiv r_t/r_s$. We take the relation between the substructure mass and its scale radius to be (see Appendix A)

$$\left(\frac{r_s}{1 \text{ kpc}} \right) = (1.0 \pm 0.3) \left(\frac{M_{\text{sub}}}{10^9 M_\odot} \right)^{0.735}, \quad (15)$$

where we have taken into account the scatter in this relation as inferred by N -body simulations [124]. We also take the tidal truncation radius to obey the standard relation [125,126]

$$r_t = \left(\frac{M_{\text{sub}}}{[2 - d \ln M_{\text{main}}/d \ln r_{3D}] M_{\text{main}}(< r_{3D})} \right)^{1/3} r_{3D}, \quad (16)$$

where r_{3D} is the three-dimensional distance between the mass substructure and the center of the main lens halo and $M_{\text{main}}(< r_{3D})$ is the fraction of the mass of the main halo contained in a sphere of radius r_{3D} . For a spherical main lens with a power-law convergence profile

$$\kappa_{\text{main}}(r) = \frac{1}{2} \left(\frac{b}{r} \right)^{2-\alpha_{\text{main}}}, \quad (\alpha_{\text{main}} \neq 2) \quad (17)$$

where b is a length scale closely related to the Einstein radius of the main lens, r is the projected two-dimensional distance from the center of the lens, and α_{main} is the power-law index of the density profile, the tidal truncation radius takes the form (see Appendix A for more details)

$$r_t = \left(\frac{\alpha_{\text{main}}}{2 - \alpha_{\text{main}}} \frac{\Gamma(\frac{2-\alpha_{\text{main}}}{2})}{\Gamma(\frac{3-\alpha_{\text{main}}}{2})} \frac{M_{\text{sub}}}{2\sqrt{\pi}\Sigma_{\text{crit}}b^2} \right)^{1/3} \left(\frac{b}{r_{3D}} \right)^{\frac{\alpha_{\text{main}}}{3}} r_{3D}, \quad (\alpha_{\text{main}} \neq 2), \quad (18)$$

where $\Gamma(x)$ is the gamma function and where Σ_{crit} is the critical density for lensing. The substructures are taken to be spatially distributed in two-dimensional projections according to a “cored” profile for $0 < r < R_{\text{max}}$ given by

$$\mathcal{P}_r(r, \theta) = \left(\frac{1}{2\pi r_c^2} \frac{1}{W(R_{\text{max}}/r_c) - 1} \right) \frac{1}{(1 + (r/r_c))^2},$$

$$\text{where } W(x) = \frac{1}{1+x} + \ln(1+x), \quad (19)$$

and where r_c is the core radius of the substructure distribution. This spatial distribution profile constitutes a good approximation to the radial substructure distribution found in N -body simulations [106]. The core radius r_c is found to be a large fraction of the main halo virial radius. Here, we take $r_c = 30R_{\text{ein}}$, where R_{ein} is the Einstein radius of the smooth lens. For a typical galaxy-scale gravitational lens with $R_{\text{ein}} \sim 1''$, this gives $r_c \sim 189$ kpc for a lens at redshift $z_{\text{lens}} \sim 0.5$. We define the boundary between the local and distributed population to lie at $R_{\text{min}} = 3R_{\text{ein}}$ and also choose $R_{\text{max}} = 65R_{\text{ein}}$. We note that r_{3D} is related to r via $r_{3D} = \sqrt{r^2 + h^2}$, where h is the projection of r_{3D} along the line of sight, which must lie in the range $-\sqrt{R_{\text{max}}^2 - r^2} \leq h \leq \sqrt{R_{\text{max}}^2 - r^2}$ for a spherical halo. When we generate the Monte Carlo realization, we first choose r from the probability distribution in Eq. (19), and then randomly pick h from the above range in order to generate the value of r_{3D} . We note though that in a realistic halo, the values of h will not in general be *uniformly* distributed within the above range. However, since h only enters in the calculation of the truncation radius, the impact of this approximation on our results is very small.

We take the substructure to be distributed in mass according to a power-law mass probability distribution

$$\mathcal{P}_M(M_{\text{sub}}) \equiv \frac{1}{N} \frac{dN}{dM_{\text{sub}}} = \frac{(\beta + 1)M_{\text{sub}}^\beta}{M_{\text{high}}^{\beta+1} - M_{\text{low}}^{\beta+1}}, \quad M_{\text{low}} < M_{\text{sub}} < M_{\text{high}}, \quad (\beta \neq -1), \quad (20)$$

where β is the power law index, and where M_{high} and M_{low} are the highest and lowest subhalo masses inside the lens galaxy, respectively. As was found numerically in Ref. [106], we take $\beta = -1.9$. We also choose $M_{\text{high}} = 10^{10} M_\odot$ and $M_{\text{low}} = 10^7 M_\odot$. While M_{low} is typically much lower in standard cold dark matter models [127,128], this latter choice ensures that the number of mass substructures inside the lens galaxy is manageable within our Monte Carlo realizations. The actual number of mass

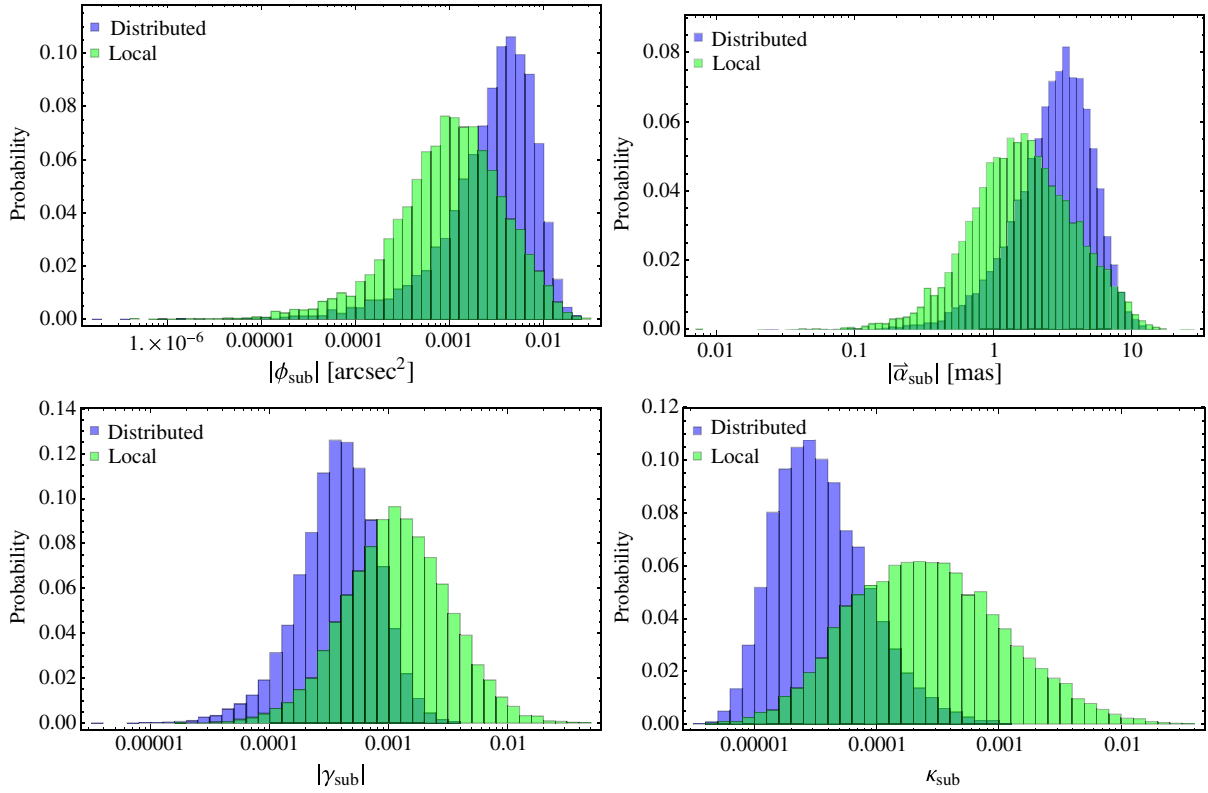


FIG. 2. Probability distributions for the local and distributed contributions to the linear lensing quantities ϕ_{sub} , α_{sub} , γ_{sub} , and κ_{sub} . These quantities are evaluated at the Einstein radius of the main lens, which we take to be $R_{\text{ein}} = 1''$. We assume the lens to be at redshift $z_{\text{lens}} = 0.5$ with a source at redshift $z_{\text{source}} = 1$, yielding a critical density for lensing $\Sigma_{\text{crit}} = 1.19 \times 10^{11} M_{\odot}/\text{arcsec}^2$. We define the divide between the distributed and local contributions to lie at $R_{\text{min}} = 3R_{\text{ein}}$ and include mass substructures up to $R_{\text{max}} = 65R_{\text{ein}}$. The substructures are spatially distributed according to a cored profile [Eq. (19)] with core radius $r_c = 30R_{\text{ein}}$. Mass substructures are taken to have a smoothly truncated NFW profile with tidal truncation radius that depends on subhalo mass and halo-centric distance as given in Eqs. (16) and (18). We use a subhalo mass-concentration relation derived from N -body simulations [124] and also implement the scatter about this relation [Eq. (15)]. We assume a power law subhalo mass function with slope $\beta = -1.9$ between $M_{\text{low}} = 10^7 M_{\odot}$ and $M_{\text{high}} = 10^{10} M_{\odot}$. We take the average lensing convergence in mass substructure at the Einstein radius to be $\langle \kappa_{\text{sub}}(R_{\text{ein}}) \rangle = 0.001$.

substructures indeed the lens galaxy is taken to be Poisson distributed around a mean value given by

$$\langle N \rangle = \frac{\langle \kappa_{\text{sub}}(R_{\text{ein}}) \rangle}{\int dM_{\text{sub}} \int r dr d\theta \mathcal{P}_M(M_{\text{sub}}) \mathcal{P}_r(r, \theta) \kappa_{\text{tNFW}}(|\mathbf{r} - \mathbf{R}_{\text{ein}}|)}, \quad (21)$$

where the angular bracket denotes ensemble averaging over many substructure realizations of the lens halo and $\kappa_{\text{tNFW}}(\mathbf{r})$ is the convergence profile of a single smoothly truncated NFW subhalo as given in Ref. [123]. Equation (21) follows from the independence of subhalos within the lens galaxy. We take the average convergence in mass substructures at the Einstein radius of the main lens to be $\langle \kappa_{\text{sub}}(R_{\text{ein}}) \rangle = 0.001$. We note that setting $\langle \kappa_{\text{sub}} \rangle$ as above is equivalent to choosing an overall normalization for the subhalo mass function [see Eq. (41) below for more details].

We illustrate in Fig. 2 the probability distributions for both the distributed and local contributions to the total

linear lensing quantities for our 10^4 Monte Carlo realizations. For lensing potential and deflection fluctuations, we observe that the most *probable* fluctuations are dominated by the distributed substructures. This result can be explained by a simple geometrical argument. Indeed, while the contribution to the net deflection from substructures inside a thin ring of radius r decays as $1/r$ for increasing r , the number of mass substructures inside the thin ring grows as r for $r < r_c$. Thus, inside the core radius of the substructure distribution, mass substructures located in a distant ring can contribute just as much to the total deflection as substructures much closer to the lensed images. A similar argument applies to lensing potential fluctuations. This indicates that detailed analyses of time delays and astrometric fluctuations caused by mass substructures can yield important information about the distributed population of satellite surrounding lens galaxies.

On the other hand, the substructure contribution to the shear and convergence (which determine the magnification perturbations) at a typical image position is largely

dominated by the local subhalos. This is due to the fact that shear perturbations decay as r^{-2} , while convergence fluctuations decay even faster (r^{-4} for our choice of truncated NFW profile). It implies that the contribution from distant rings of substructures is always subdominant compared to the local contribution, although the collective shear perturbations from the distributed substructures is not entirely negligible. Furthermore, for deflections, shears, and convergence, the largest fluctuations are always dominated by the local contributions. These large local perturbations, often caused by a single substructure close to a lensed image, have been used to detect individual mass substructures [20,21,36]. What Fig. 2 is showing however is that by combining magnification information (largely sensitive to γ_{sub} and κ_{sub}) with astrometric (sensitive to α_{sub}) and time delay (sensitive to ϕ_{sub}) measurements, one could infer important properties about *both* the local and distributed substructure populations inside lens galaxies. This highlights the importance of developing a unified framework to jointly handle the different lensing observables, which is a major goal of this work.

Splitting the mass substructures into two subpopulations allows us to factorize the characteristic function $Q_{\langle N \rangle}$ given in Eq. (9) as

$$Q_{\langle N \rangle}(\mathbf{k}_L | \mathbf{q}, \mathbf{q}_{\text{sub}}) = Q_{\langle N_l \rangle}^{\text{local}}(\mathbf{k}_L | \mathbf{q}, \mathbf{q}_{\text{sub}}) Q_{\langle N_d \rangle}^{\text{dist}}(\mathbf{k}_L | \mathbf{q}, \mathbf{q}_{\text{sub}}), \quad (22)$$

which follows from Eq. (12) and the independence of each substructure. Here, $\langle N \rangle = \langle N_l \rangle + \langle N_d \rangle$, where $\langle N_l \rangle$ is the average number of substructures in the local population, and where $\langle N_d \rangle$ is the average number of substructures in the distributed population. We note that we generically have $\langle N_d \rangle \gg \langle N_l \rangle$. In terms of the characteristic function for a single subhalo, this implies

$$q_1(\mathbf{k}_L | \mathbf{q}, \mathbf{q}_{\text{sub}}) = \frac{\langle N_l \rangle q_1^{\text{local}}(\mathbf{k}_L | \mathbf{q}, \mathbf{q}_{\text{sub}}) + \langle N_d \rangle q_1^{\text{dist}}(\mathbf{k}_L | \mathbf{q}, \mathbf{q}_{\text{sub}})}{\langle N \rangle}. \quad (23)$$

We can therefore separately compute the characteristic function for the local and distributed subpopulation and then combine them according to Eq. (23) to compute the overall characteristic function of linear lensing quantities. In this work, we focus on statistically characterizing the distributed population of mass substructures inside typical lens galaxies, which is the dominant contribution for the projected lensing potential and deflections. We leave to future work the characterization of the local substructure population, but we note that gravitational imaging techniques [21,36,43,87] and resolved spectroscopy [24,72,88] can provide information about certain regions of the local substructure population.

B. Distributed substructure analysis for potential and deflection perturbations

In this section, we outline our calculations of the characteristic function $q_1^{\text{dist}}(\mathbf{k}_L | \mathbf{q}, \mathbf{q}_{\text{sub}})$ for the distributed population of mass substructures. We focus exclusively on computing the characteristic function for the projected lensing potential and the deflection perturbations since the contribution to shear and convergence perturbations from the distributed population of mass substructures is subdominant. As described above, there are key simplifying facts for the distributed substructure population:

- (i) Their overall impact on the lensing observables is small.
- (ii) We can approximate them as a collection of point masses.

In the point-mass approximation, a single subhalo can be described by three parameters: its total mass M_{sub} and its radial and angular position in the lens plane. In the notation from Sec. II, this implies $\mathbf{c}_{\text{sub}}^{(1)} = \{M_{\text{sub}}, r, \theta\}$. In order to construct the characteristic function, we need to specify the probability density function $P_{\text{sub}}(M_{\text{sub}}, r, \theta; \mathbf{q}_{\text{sub}})$ for these parameters. As in our Monte Carlo examples of Sec. III A, we assume that this density function is separable into the product of the subhalo mass function with a spatial density distribution

$$P_{\text{sub}}(M_{\text{sub}}, r, \theta; \mathbf{q}_{\text{sub}}) = \mathcal{P}_M(M_{\text{sub}}; \mathbf{q}_{\text{sub}}) \mathcal{P}_r(r, \theta; \mathbf{q}_{\text{sub}}). \quad (24)$$

For cold dark matter, this separability is supported by N -body simulations over a wide range of subhalo masses [106,129,130]. It remains to be seen whether this separability holds for more general dark matter models or when baryonic feedback is taken into account. In the cases for which the mass and spatial distributions are not separable, one could split the subhalo population into several subpopulations that each have with their own spatial distribution. For simplicity, we assume here that Eq. (24) is valid, but it is clear that our analysis could also be carried out without this assumption. As we see below, we do not need to specify an explicit form for the subhalo mass function and position distribution at this point since all important quantities can be expressed as statistical moments of these distributions.

Before going through the detailed derivation of $q_1^{\text{dist}}(\mathbf{k}_L | \mathbf{q}, \mathbf{q}_{\text{sub}})$, it is informative to heuristically derive what we expect the answer to be. As we discussed in the previous subsection, potential and deflection perturbations obtain contributions from a broad spatial projected area surrounding the strong lensing region. The resulting large number of mass substructures contributing to the total potential and deflection perturbations implies that the central limit theorem is applicable, and we thus expect Gaussian statistics to be approximately valid for these

linear lensing quantities. In this approximation, the statistical properties of the linear lensing quantities are entirely specified by a covariance matrix \mathbf{C}_{sub} with general scaling given by

$$\mathbf{C}_{\text{sub}}^{ij} \propto \langle N_d \rangle \left[\int dm \mathcal{P}_m(m; \mathbf{q}_{\text{sub}}) m^2 \right] \times \left[\int d^2 \mathbf{r} \mathcal{P}_r(r, \theta; \mathbf{q}_{\text{sub}}) \mathcal{O}_L^i \mathcal{O}_L^j \right], \quad (25)$$

where the leading factor arises since the variance of the sum of $\langle N_d \rangle$ normal random variables is $\langle N_d \rangle$ times the variance of a single normal random variable. The second factor arises because the linear lensing quantities are always proportional to the subhalo mass, and the third factor is the spatial two-point function of the linear lensing quantities. Here, \mathcal{O}_L^i stands for the spatial dependence of the i th linear lensing quantity. As we see below, this scaling comes out naturally of our analysis.

We now turn our attention to the detailed derivation of the above scaling as well as the leading order deviations from the Gaussian approximation. The lensing potential difference $\phi_{\text{sub}}(\mathbf{x}_i)$ between an image position \mathbf{x}_i and a reference point \mathbf{x}_{ref} caused by a point mass M at position \mathbf{x} is given by

$$\phi_{\text{sub}}(\mathbf{x}_i) = m \ln \left[\frac{|\mathbf{x}_i - \mathbf{x}|}{|\mathbf{x}_{\text{ref}} - \mathbf{x}|} \right], \quad (26)$$

where $m \equiv M_{\text{sub}}/(\pi \Sigma_{\text{crit}})$. Since Σ_{crit} is the critical mass density for lensing, m has units of area. Since $|\mathbf{x}_i| \sim R_{\text{ein}} \ll |\mathbf{x}|$ for a typical distributed dark matter substructure, we can write down the lensing potential difference at an image location as a multipole expansion. Converting to polar coordinates with $\mathbf{x} = (r \cos \theta, r \sin \theta)$, we obtain

$$\vec{A}_p = \left(\frac{r_i^p \cos \{p\theta_i\} - r_{\text{ref}}^p \cos \{p\theta_{\text{ref}}\}}{p}, r_i^{p-1} \cos \{(p-1)\theta_i\}, -r_i^{p-1} \sin \{(p-1)\theta_i\} \right), \quad (30)$$

$$\vec{B}_p = \left(\frac{r_i^p \sin \{p\theta_i\} - r_{\text{ref}}^p \sin \{p\theta_{\text{ref}}\}}{p}, r_i^{p-1} \sin \{(p-1)\theta_i\}, r_i^{p-1} \cos \{(p-1)\theta_i\} \right), \quad (31)$$

where $p \geq 1$. We emphasize that \vec{A}_p and \vec{B}_p are constant vectors that only depend on the configuration of lensed images and are thus independent of the mass substructure population. Taking \mathbf{k}_L to be the Fourier conjugate of the stochastic vector $\vec{\mathcal{O}}_L$, the characteristic function for a single dark matter substructure can be written as

$$\begin{aligned} q_1^{\text{dist}}(\mathbf{k}_L | \mathbf{q}, \mathbf{q}_{\text{sub}}) &= \int_{\mathcal{H}_d} d^2 r \int dm e^{i \mathbf{k}_L \cdot \vec{\mathcal{O}}_L} P_{\text{sub}}(m, r, \theta; \mathbf{q}_{\text{sub}}) \\ &= 1 + \int_{\mathcal{H}_d} d^2 r \int dm (e^{i \mathbf{k}_L \cdot \vec{\mathcal{O}}_L} - 1) P_{\text{sub}}(m, r, \theta; \mathbf{q}_{\text{sub}}), \end{aligned} \quad (32)$$

where \mathcal{H}_d denotes the area of the distributed domain of the lens halo and where we pulled out the leading factor of unity since we are only interested in the difference $q_1(\mathbf{k}_L | \mathbf{q}, \mathbf{q}_{\text{sub}}) - 1$. Evaluating the above integrals is the most difficult part of

$$\phi_{\text{sub}}(\mathbf{x}_i) = -m \sum_{p=1}^{\infty} \frac{1}{r^p} [A_p(\mathbf{x}_i) \cos(p\theta) + B_p(\mathbf{x}_i) \sin(p\theta)], \quad (27)$$

where the dimensionless series coefficients are

$$\begin{aligned} A_p(\mathbf{x}_i) &= \frac{1}{p} (r_i^p \cos(p\theta_i) - r_{\text{ref}}^p \cos(p\theta_{\text{ref}})), \\ B_p(\mathbf{x}_i) &= \frac{1}{p} (r_i^p \sin(p\theta_i) - r_{\text{ref}}^p \sin(p\theta_{\text{ref}})), \end{aligned} \quad (28)$$

where we have used $\mathbf{x}_i = (r_i \cos \theta_i, r_i \sin \theta_i)$ and $\mathbf{x}_{\text{ref}} = (r_{\text{ref}} \cos \theta_{\text{ref}}, r_{\text{ref}} \sin \theta_{\text{ref}})$. Since the deflections are simply related to the lensing potential by derivatives, that is, $\vec{\alpha}_{\text{sub}}(\mathbf{x}_i) = \vec{\nabla}_{\mathbf{x}_i} \phi_{\text{sub}}$, we can write expansions similar to Eq. (27) for each of these quantities. The only difference is that the series coefficients for $\vec{\alpha}_{\text{sub}}$ are derivatives of $A_p(\mathbf{x}_i)$ and $B_p(\mathbf{x}_i)$. Taking $\vec{\mathcal{O}}_L \equiv \Delta \mathbf{t}_L / m$ to denote the vector containing all the stochastic random variables describing the perturbations to the linear lensing quantities, we can thus write

$$\vec{\mathcal{O}}_L = - \sum_{p=1}^{\infty} \frac{1}{r^p} [\vec{A}_p \cos(p\theta) + \vec{B}_p \sin(p\theta)]. \quad (29)$$

We note that we have divided out the leading factor of the subhalo mass in the above definition since it only leads to an overall rescaling of $\vec{\mathcal{O}}_L$. In general, $\vec{\mathcal{O}}_L$ would contain the stochastic variables $\phi_{\text{sub}}^{(i)}$ and $\vec{\alpha}_{\text{sub}}^{(i)}$ for the lensing potentials and deflections, respectively, evaluated at all possible image positions $i \in N_{\text{img}}$. For instance, in the case of a single image with label i , we have $\vec{\mathcal{O}}_L^{(i)} = \frac{1}{m} \{\phi_{\text{sub}}^{(i)}, \alpha_{\text{sub},x}^{(i)}, \alpha_{\text{sub},y}^{(i)}\}$ and

the calculation. Clearly, for $|\mathbf{k}_L| \ll 1/(m|\vec{\mathcal{O}}_L|)$, the phase factor is nearly equal to unity and $q_1 \rightarrow 1$, while $q_1 \rightarrow 0$ for $|\mathbf{k}_L| \gg 1/(m|\vec{\mathcal{O}}_L|)$ since the phase is rapidly oscillating in this regime. We expand $q_1(\mathbf{k}_L|\mathbf{q}, \mathbf{q}_{\text{sub}})$ in a power series of mass and spatial moments

$$q_1^{\text{dist}}(\mathbf{k}_L|\mathbf{q}, \mathbf{q}_{\text{sub}}) = 1 + \sum_{n=1}^{\infty} \frac{i^n \langle m^n \rangle}{n!} \int_{\mathcal{H}_d} d^2r \mathcal{P}_r(r, \theta; \mathbf{q}_{\text{sub}}) (\mathbf{k}_L \cdot \vec{\mathcal{O}}_L)^n, \quad (33)$$

where the mass moments are given by

$$\langle m^n \rangle \equiv \int dm \mathcal{P}_m(m; \mathbf{q}_{\text{sub}}) m^n. \quad (34)$$

For conciseness, the simplification of the spatial integral appearing in Eq. (33) is presented in Appendix B. After these simplifications, the characteristic function for the linear quantities $\vec{\mathcal{O}}_L$ in the presence of a single mass substructure then takes the form

$$q_1^{\text{dist}}(\mathbf{k}_L|\mathbf{q}, \mathbf{q}_{\text{sub}}) = 1 + \sum_{n=1}^{\infty} \frac{(-1)^n \langle m^n \rangle}{n!} \left(\sum_{\|\mathbf{p}\|=n} \binom{n}{\mathbf{p}} K_{\mathbf{p}} \prod_{t=1}^{N_{\text{mult}}} (i\mathbf{k}_L \cdot \vec{A}_t)^{p_t} \prod_{s=1}^{N_{\text{mult}}} (i\mathbf{k}_L \cdot \vec{B}_s)^{p_{N_{\text{mult}}+s}} \right), \quad (35)$$

where $\mathbf{p} = \{p_1, p_2, \dots, p_{2N_{\text{mult}}}\}$ is a multi-index with $\|\mathbf{p}\| = \sum_{j=1}^{2N_{\text{mult}}} p_j$, and where the kernel $K_{\mathbf{p}}$ is given by Eq. (B2). It is understood that if a given \vec{A}_t or \vec{B}_t vanishes, then the corresponding p_t must also vanish. We emphasize that the kernel $K_{\mathbf{p}}$ encodes all the information about the spatial distribution of mass substructures within the lens halo. This kernel can be computed for any halo geometry and mass substructure distribution. Finally, we can use Eq. (9) to compute the characteristic function in the presence of a whole population of mass substructures

$$\mathcal{Q}_{\langle N_d \rangle}^{\text{dist}}(\mathbf{k}_L|\mathbf{q}, \mathbf{q}_{\text{sub}}) = \exp \left[\langle N_d \rangle \sum_{n=1}^{\infty} \frac{(-1)^n \langle m^n \rangle}{n!} \left(\sum_{\|\mathbf{p}\|=n} \binom{n}{\mathbf{p}} K_{\mathbf{p}} \prod_{t=1}^{N_{\text{mult}}} (i\mathbf{k}_L \cdot \vec{A}_t)^{p_t} \prod_{s=1}^{N_{\text{mult}}} (i\mathbf{k}_L \cdot \vec{B}_s)^{p_{N_{\text{mult}}+s}} \right) \right]. \quad (36)$$

At leading order, this characteristic function has a Gaussian behavior,

$$\mathcal{Q}_{\langle N_d \rangle}^{\text{dist}}(\mathbf{k}_L) \propto e^{i\mathbf{u} \cdot \mathbf{k}_L - \frac{1}{2} \mathbf{k}_L^T \mathbf{C}_{\text{sub}} \mathbf{k}_L}, \quad (37)$$

where $\mathbf{u} \equiv \langle \Delta \mathbf{t}_L \rangle$ is the mean vector and $\mathbf{C}_{\text{sub}} \equiv \langle \Delta \mathbf{t}_L \Delta \mathbf{t}_L^T \rangle$ is the covariance matrix. We note that in the case of circular symmetry of the galactic halo, the mean vector \mathbf{u} exactly vanishes. We give in Appendix D some useful expressions for the covariance matrix in the case of circular symmetry for two different spatial distributions.

The non-Gaussian terms in Eq. (36) essentially forms a multivariate Edgeworth expansion (see, e.g., Refs. [102,131]) with successive term decaying as $\langle N_d \rangle^{1-n/2}$. We show the details of this expansion in Appendix C, but it is instructive to consider the magnitude of the non-Gaussian contributions to $\mathcal{Q}_{\langle N_d \rangle}^{\text{dist}}(\mathbf{k}_L)$ in order to assess the validity of the leading Gaussian approximation. At each order n in the $1/\langle N_d \rangle^{n/2-1}$ expansion, the leading order non-Gaussian contribution takes the general form

$$\frac{1}{n! \langle N_d \rangle^{n/2-1}} \frac{\langle m^n \rangle}{\langle m^2 \rangle^{n/2}} \frac{\langle \mathcal{O}_L^n \rangle}{\langle \mathcal{O}_L^2 \rangle^{n/2}}, \quad (n \geq 3). \quad (38)$$

Here, we use the compact notation $\langle \mathcal{O}_L^n \rangle$ to represent all possible spatial n -point functions of the different linear lensing quantities. In order to evaluate the above expression, we need to specify the mass function and spatial distribution of mass substructures. For illustration, we take the spatial distribution given in Eq. (19) and write the mass function as

$$\frac{dN}{dM_{\text{sub}}} = a_0 \left(\frac{M_{\text{sub}}}{M_0} \right)^{\beta}, \quad (39)$$

where a_0 is the mass function normalization and M_0 is a reference mass scale. Using Eq. (21), the expected number of mass substructures in the distributed region is then

$$\langle N_d \rangle = \frac{a_0}{M_0^{\beta}} \frac{M_{\text{high}}^{\beta+1} - M_{\text{low}}^{\beta+1}}{\beta + 1} (1 - \mathcal{P}_r(<R_{\text{min}})), \quad (\beta \neq -1), \quad (40)$$

where $\mathcal{P}_r(<R_{\text{min}})$ is the cumulative probability of finding a mass substructure within a disk of radius R_{min} . We take $M_0 = M_{\text{high}}$ throughout this work. We note that in the point-mass limit, the convergence in mass substructures is related to the mass function given in Eq. (39) via

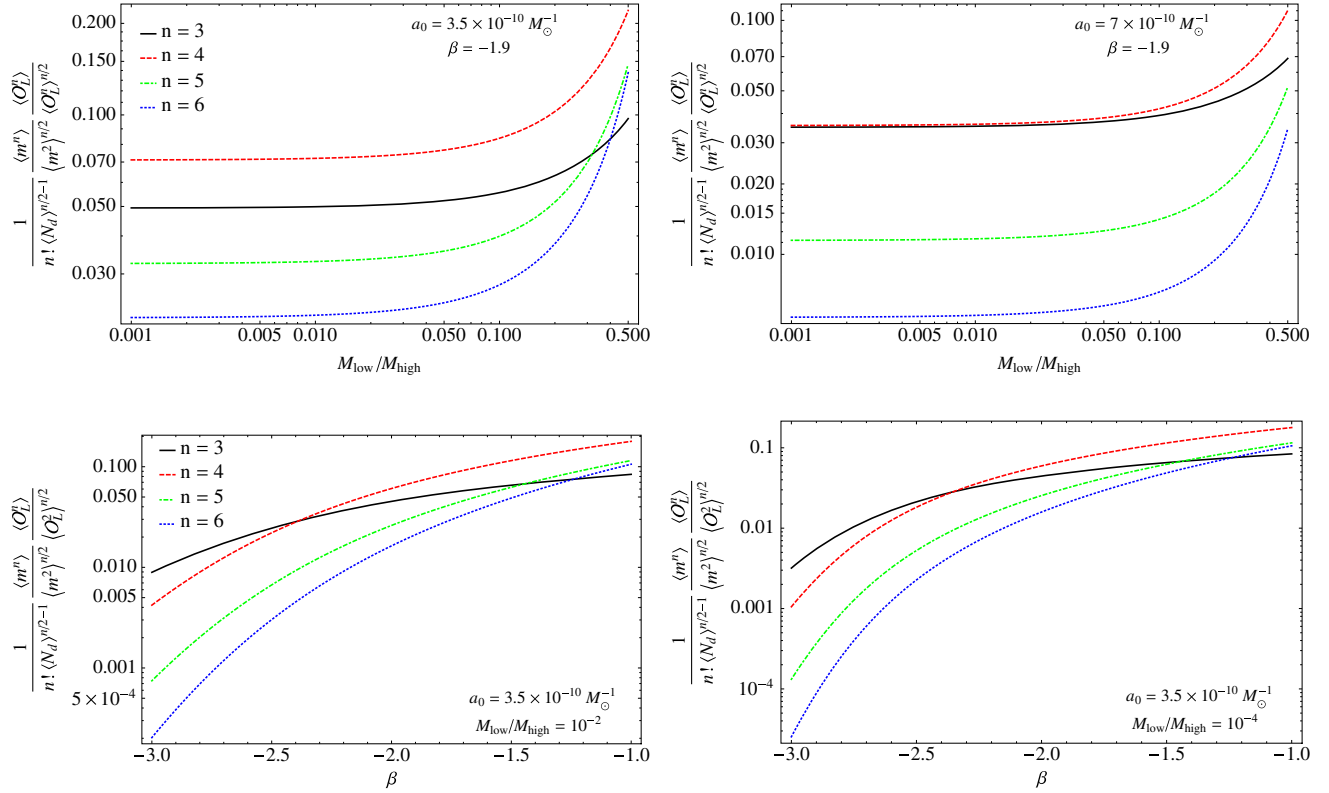


FIG. 3. Non-Gaussian contributions to the Edgeworth expansion of the characteristic function $Q_{(N_d)}^{\text{dist}}(\mathbf{k}_L)$ for different values of the mass function parameters. These curves characterize the degree of non-Gaussianity of the probability distribution of the linear lensing quantities. A value of unity on these plots indicate an $O(1)$ deviation from Gaussianity. We assume a power law mass function as given in Eq. (39) with $M_{\text{high}} = 10^{10} M_{\odot}$ and also take the distributed mass substructures to be spatially located between $R_{\text{min}} = 3R_{\text{ein}}$ and $R_{\text{max}} = 65R_{\text{ein}}$ according to Eq. (19) with $r_c = 30R_{\text{ein}}$. We illustrate the leading contribution at each order n for $n = 3$ to $n = 6$. Here, $\langle N_d \rangle$ is computed as in Eq. (40). The spatial moments $\langle O_L^n \rangle$ are computed assuming that O_L is a deflection at a single image position, but similar results would be obtained for the lensing potential. Each panel illustrates different mass function parameters as indicated. The top panels fix $\beta = -1.9$ and display the dependence of the non-Gaussian corrections on the ratio $M_{\text{low}}/M_{\text{high}}$ for two different values of a_0 . In the bottom panels, we fix a_0 and display the dependence on the mass function slope for two values of $M_{\text{low}}/M_{\text{high}}$.

$$\langle \kappa_{\text{sub}}(r_{\text{ref}}) \rangle = \frac{a_0}{M_0^\beta \Sigma_{\text{crit}}} \frac{M_{\text{high}}^{\beta+2} - M_{\text{low}}^{\beta+2}}{\beta + 2} \mathcal{P}_r(r_{\text{ref}}), \quad (\beta \neq -2), \quad (41)$$

where r_{ref} is a reference radius (e.g., R_{ein}) where the convergence is evaluated.

We illustrate in Fig. 3 the non-Gaussian contributions given in Eq. (38) evaluated from $n = 3$ to $n = 6$ for different mass function parameters. Here, we take O_L to represent a lensing deflection, but similar results would be obtained for the lensing potential. The upper panels illustrate the dependence of the non-Gaussian contributions on the lowest subhalo mass for two different values of the mass function normalization a_0 with $\beta = -1.9$. For comparison, pure cold dark matter simulations yield $a_0 \sim 3.8 \times 10^{-10} M_{\odot}^{-1}$ at the pivot point $M_0 = 10^{10} M_{\odot}$ with $\beta = -1.9$ [106]. We observe that for $M_{\text{low}}/M_{\text{high}} \lesssim 0.1$ the non-Gaussian contributions are subdominant for the fiducial values of $a_0 = 3.5 \times 10^{-10} M_{\odot}^{-1}$ and $\beta = -1.9$.

Interestingly, the largest non-Gaussian contribution comes from the $n = 4$ term, which implies that the probability density function of linear lensing quantities will primarily pick up a nonzero excess kurtosis in this case. Further increasing the normalization of the subhalo mass function suppresses non-Gaussianities even more (upper right panel) since $\langle N_d \rangle \propto a_0$. However, as $M_{\text{low}}/M_{\text{high}} \rightarrow 1$, the non-Gaussian contributions rapidly rise since the mass substructure population becomes dominated by a very limited number of massive subhalos and the applicability of the central limit theorem wanes.

The lower panels of Fig. 3 display the dependence of the non-Gaussian corrections on the slope of the substructure mass function. Here, we fix the ratio $M_{\text{low}}/M_{\text{high}}$ and the amplitude of the mass function at $M_0 = M_{\text{high}}$. We observe that as β is made steeper (more negative) the non-Gaussian corrections rapidly decay since the number of mass substructures quickly rises with a steepening slope. Decreasing the ratio $M_{\text{low}}/M_{\text{high}}$ has little effect for $\beta > -2$ but does lead a faster decay of non-Gaussianities for $\beta < -2$. Again,

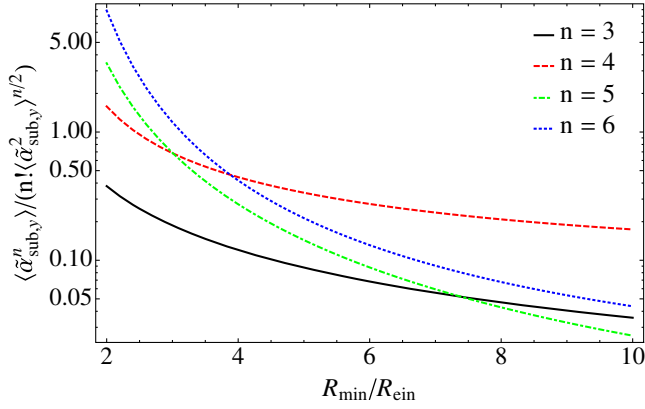


FIG. 4. Dependence of the non-Gaussian spatial moments of a deflection y-component on the value of R_{min} . We take the distributed mass substructures to be spatially distributed between R_{min} and $R_{\text{max}} = 65R_{\text{ein}}$ according to Eq. (19) with $r_c = 30R_{\text{ein}}$. It is understood here that the mass dependence of the lensing deflection has been divided out, that is, $\tilde{\alpha}_{\text{sub},y} \equiv \alpha_{\text{sub},y}/m$. We see that the non-Gaussian moments decay as R_{min} is increased.

we observe that the $n = 4$ term dominates the non-Gaussianities when the mass function slope $\beta \gtrsim -2.3$ for the realistic normalization of the mass function shown. We confirm this observation by comparing our analytical results to Monte Carlo realizations in the next subsection.

We note that we can also suppress non-Gaussianities by increasing R_{min} . Indeed, the non-Gaussian spatial moments $\langle \mathcal{O}_L^n \rangle / \langle \mathcal{O}_L^2 \rangle^{n/2}$ rapidly decreases as R_{min} is increased as shown in Fig. 4. For definiteness, we illustrate there the ratio of non-Gaussian spatial moments for a single component of a lensing deflection. The results would be very similar for other linear lensing quantities. From a practical point-of-view, we would like R_{min} to be as small as possible in order to encompass as many mass substructures as possible in the distributed analysis. On the other hand, we also need to choose a value of R_{min} large enough for the expansion of Eq. (36) to rapidly converge. Our tests show that a minimal radius in the range $3R_{\text{ein}} \lesssim R_{\text{min}} \lesssim 5R_{\text{ein}}$ generally provides a good compromise between these two criteria for the power law mass function considered in this work. Of course, for a different choice of mass function one should adjust R_{min} in order to insure the convergence of the Edgeworth expansion.

The picture that emerges from the considerations above is that for the CDM-relevant case of $\beta \sim -1.9$, $M_{\text{low}}/M_{\text{high}} \ll 1$, and a realistic normalization of the substructure mass function supported by simulations, the non-Gaussian contributions to Eq. (36) are subdominant, and the joint probability density function of linear lensing quantities will thus be well approximated by a multivariate Gaussian. In this physically relevant region, a useful criterion for the validity of the Gaussian approximation is

$$a_0 \gtrsim \frac{10}{M_{\text{high}}} \frac{(\beta + 3)^2}{4!(\beta + 5)} \frac{\langle \mathcal{O}_L^4 \rangle}{\langle \mathcal{O}_L^2 \rangle^2}, \quad (42)$$

which is valid for $M_{\text{low}}/M_{\text{high}} \ll 1$, $\beta > -3$, and where a_0 is the amplitude of the mass function at $M_0 = M_{\text{high}}$. For our choice of spatial distribution given in Eq. (19) with $R_{\text{min}} = 3R_{\text{ein}}$, $R_{\text{max}} = 65R_{\text{ein}}$, $r_c = 30R_{\text{ein}}$, and assuming $M_{\text{high}} = 10^{10} M_{\odot}$, $\beta = -1.9$, and $R_{\text{ein}} = 1''$, this criterion reads $a_0 \gtrsim 2.3 \times 10^{-10} M_{\odot}^{-1}$ when \mathcal{O}_L is a lensing deflection. This condition would be slightly relaxed if \mathcal{O}_L is taken to be the lensing potential instead. Whenever this condition is satisfied, the characteristic function expansion given in Eq. (36) [see also Eq. (C6)] provides an accurate description of the statistical properties of perturbations to the linear lensing quantities caused by distributed mass substructures.

C. Validity of analytical approach

We now test the validity of the characteristic function-based approach by comparing its prediction to the distributions of linear lensing quantities obtained by considering a large number of Monte Carlo realizations of distributed substructure populations. As in the analytical calculation of the previous section, we treat the distributed mass substructures as point masses that are spatially distributed according to the cored profile given in Eq. (19) with $r_c = 30R_{\text{ein}}$, between $R_{\text{min}} = 3R_{\text{ein}}$ and $R_{\text{max}} = 65R_{\text{ein}}$. We consider the distribution of linear lensing quantities at two fiducial image positions located at $\mathbf{x}_1 = (0, R_{\text{ein}})$ and $\mathbf{x}_2 = (R_{\text{ein}}, 0)$, and take $R_{\text{ein}} = 1''$. For concreteness, we assume a lens at redshift $z_{\text{lens}} = 0.5$ with a source at redshift $z_{\text{src}} = 1$, which yields a critical density for lensing $\Sigma_{\text{crit}} = 1.19 \times 10^{11} M_{\odot}/\text{arcsec}^2$. To compute the final probability distribution of linear lensing quantities, we sample the characteristic function given in Eq. (C6) on a grid of \mathbf{k}_L and use a fast Fourier transform algorithm to perform the transformation back to configuration space.

In Fig. 5, we compare our analytical predictions to the results from Monte Carlo simulations of distributed substructure populations for a subhalo mass function as given in Eq. (20) with $M_{\text{low}} = 10^7 M_{\odot}$, $M_{\text{high}} = 10^{10} M_{\odot}$, and $\beta = -1.9$, with a normalization given by $\langle \kappa_{\text{sub}}(R_{\text{ein}}) \rangle = 0.001$. We display different projections of the joint probability density function for the linear lensing quantities evaluated at the two fiducial image positions. The gray points in the 2D plots and the blue histograms along the diagonal show the results from the Monte Carlo realizations of distributed substructure population. We show in solid black the results gotten by only keeping the leading Gaussian term in Eq. (C6), while the dashed red lines show the results obtained by keeping all terms up to order $O(\langle N_d \rangle^{-2})$ in the Edgeworth expansion. Since the mass function parameters listed above predict a relatively large number of mass substructures within the lens halo

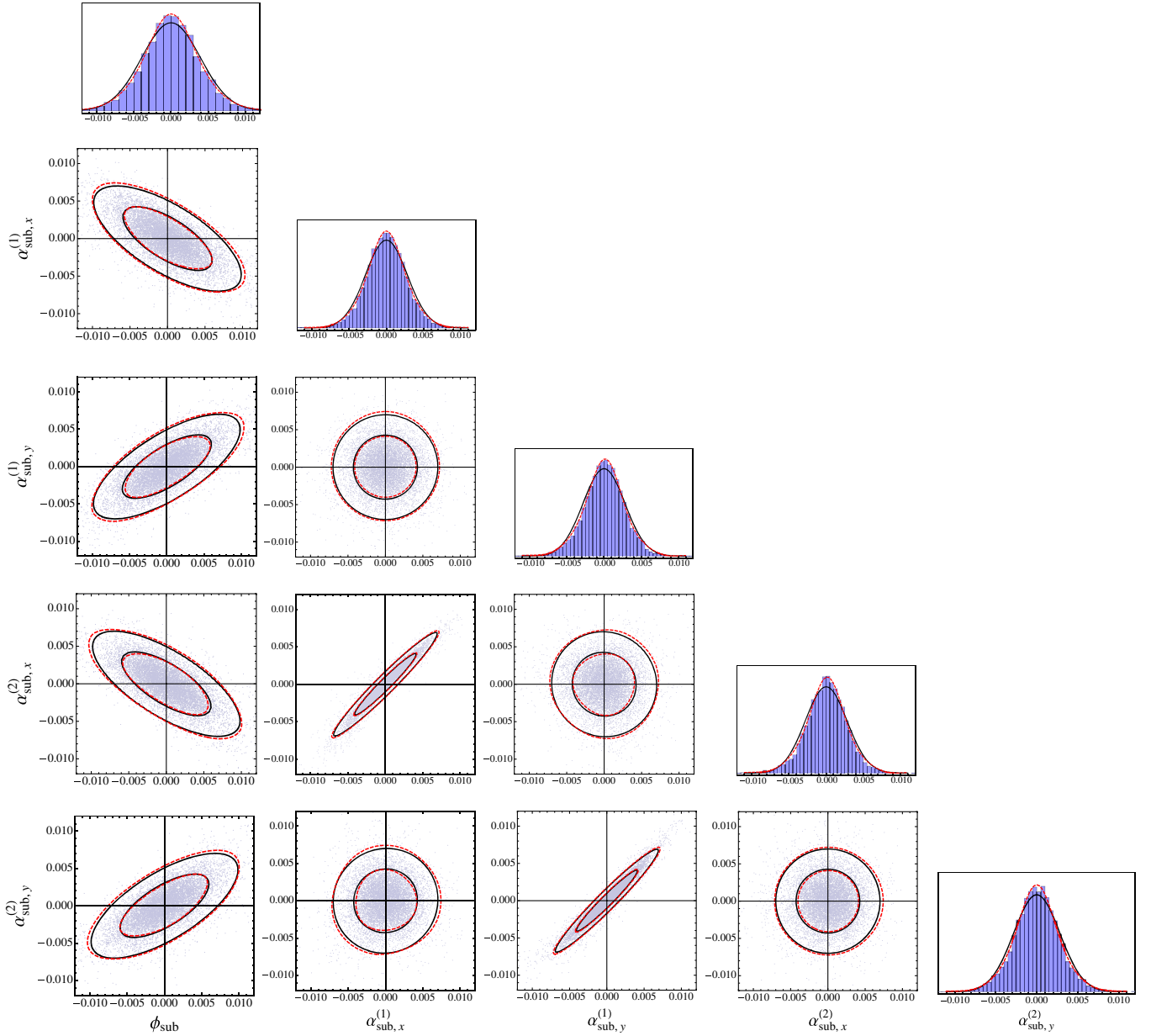


FIG. 5. Projections of the probability density function for the linear lensing quantities $\{\phi_{\text{sub}}, \alpha_{\text{sub},x}^{(1)}, \alpha_{\text{sub},y}^{(1)}, \alpha_{\text{sub},x}^{(2)}, \alpha_{\text{sub},y}^{(2)}\}$ in the presence of a distributed population of mass substructures. Here, the two images are taken to be $\mathbf{x}_1 = (0, R_{\text{ein}})$ and $\mathbf{x}_2 = (R_{\text{ein}}, 0)$, where we take $R_{\text{ein}} = 1''$. In the above, ϕ_{sub} stands for the projected potential difference between the two images. The gray points in the 2D plots and the blue histograms along the diagonal show the results from 10^4 Monte Carlo realizations of distributed point mass-like substructure population. The solid black lines display the analytical results from Sec. III B assuming a purely Gaussian characteristic function, while the dashed red lines show the results obtained by keeping all terms up to order $\langle N_d \rangle^{-2}$ in the Edgeworth expansion [Eq. (C6)]. In the 2D plots, the inner and outer contours display the 68% and 95% confidence regions, respectively. We assume the mass substructures to be spatially distributed according to Eq. (19) with $r_c = 30R_{\text{ein}}$. We also take a power law subhalo mass function with slope $\beta = -1.9$ between $M_{\text{low}} = 10^7 M_\odot$ and $M_{\text{high}} = 10^{10} M_\odot$, and take $\langle \kappa_{\text{sub}}(R_{\text{ein}}) \rangle = 0.001$. This yields an expected number of distributed mass substructures $\langle N_d \rangle = 3705$.

($\langle N_d \rangle = 3705$), the non-Gaussian contributions in Eq. (C6) are suppressed, and the overall behavior of the joint probability density function is very well captured by its leading Gaussian term. Nevertheless, we see that including the higher-order terms in the Edgeworth expansion does

improve the concordance of the analytical predictions with the Monte Carlo realizations. This is especially noticeable in the one-parameter probability densities shown along the diagonal where we observe that dashed red lines capture the nonzero excess kurtosis of the Monte Carlo realizations.

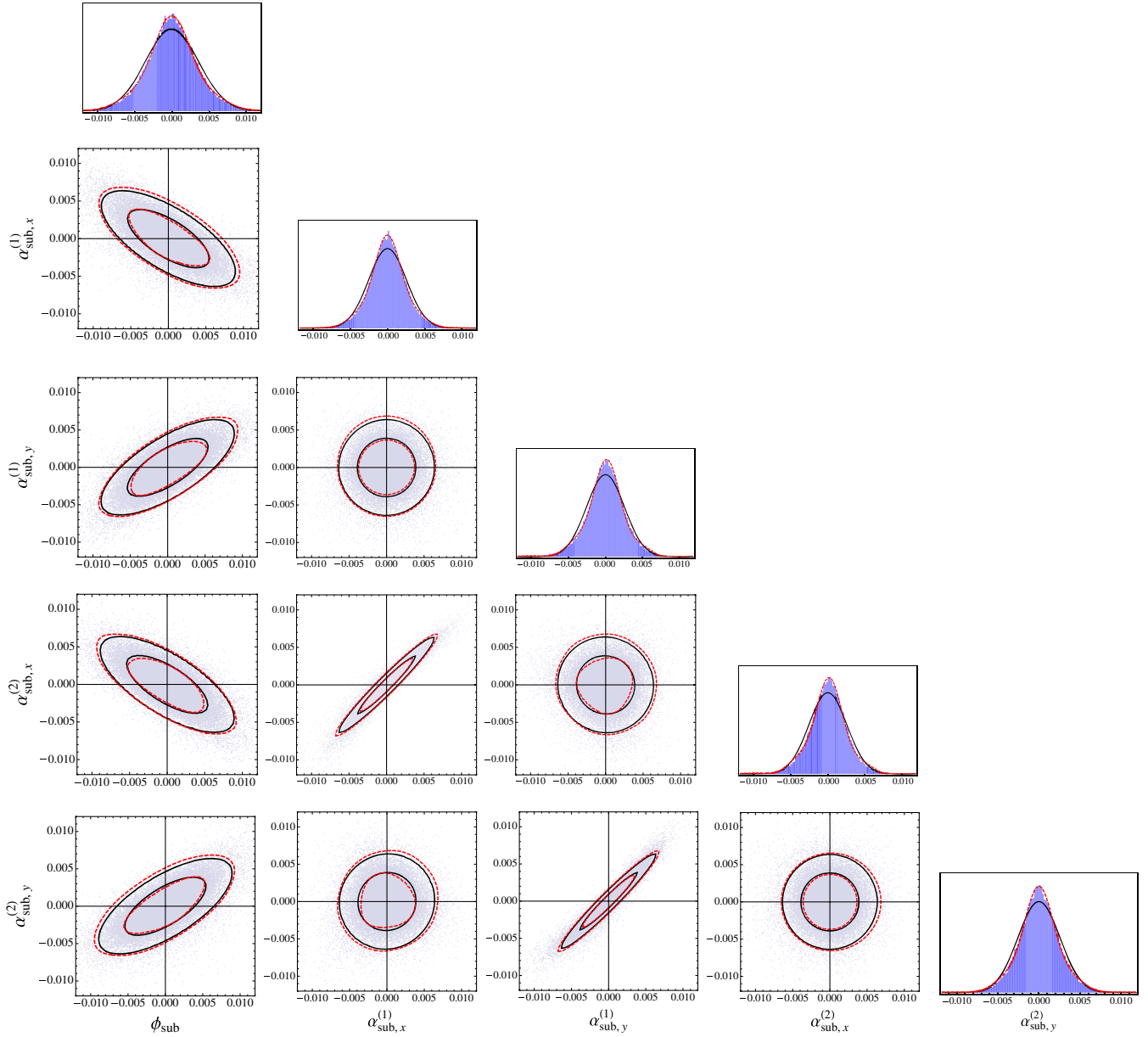


FIG. 6. Projections of the probability density function for the linear lensing quantities similar to Fig. 5 but taking $M_{\text{low}} = 2 \times 10^9 M_{\odot}$, $M_{\text{high}} = 10^{10} M_{\odot}$, and $\langle \kappa_{\text{sub}}(R_{\text{ein}}) \rangle = 3 \times 10^{-4}$. The expected number of distributed substructures is $\langle N_d \rangle = 24$. The gray points in the 2D plots and the blue histograms along the diagonal show the results from 5×10^4 Monte Carlo realizations of distributed point mass-like substructure population.

This indicates that the characteristic function expansion performed in Sec. III B does lead to the correct probability density function for the linear lensing quantities.

In Fig. 6, we display similar projections of the probability density function of linear lensing quantities, but here we choose a high value of the low mass cutoff $M_{\text{low}} = 2 \times 10^9 M_{\odot}$, together with $M_{\text{high}} = 10^{10} M_{\odot}$ and $\langle \kappa_{\text{sub}}(R_{\text{ein}}) \rangle = 3 \times 10^{-4}$. This is an example of a model with very few distributed substructures ($\langle N_d \rangle = 24$) for which the leading Gaussian approximation still works reasonably well. As clearly shown in the 1D histograms

along the diagonal of the plot, this model does have a significant excess kurtosis which is well captured by the Edgeworth expansion. Again, this highlights the usefulness of the expansion given in Eq. (C6) to understand the leading departure from Gaussianity.

Interestingly, we observe in Fig. 5 that the perturbations to linear lensing quantities from the distributed mass substructures sometimes display strong correlations among themselves. This indicates that the correlation length of perturbations to the linear lensing quantities caused by distributed substructures is larger than the typical image

separation in lens systems, consistent with the findings of Ref. [50]. More precisely, this implies that the linear lensing quantity perturbations from distributed substructures are dominated by the dipole ($p = 1$) term in the multipole expansion of Eq. (29). For deflections, this term is independent of image position which explains the very tight correlation between $\alpha_{\text{sub},x}^{(1)}$ and $\alpha_{\text{sub},x}^{(2)}$ and between $\alpha_{\text{sub},y}^{(1)}$ and $\alpha_{\text{sub},y}^{(2)}$. The scatter about this almost perfect correlation is due to the contributions from higher multipoles. We note that this scatter tends to increase at large deflection values since these are caused by substructures that are closer to the images and are thus described by higher multipoles. The correlation between deflections and lensing potential perturbations also suggests that time-delay fluctuations caused by mass substructures are usually accompanied by a corresponding shift to the image position.

Ultimately, the correlations between linear lensing quantities at different image positions are particularly interesting since it may allow one to distinguish the effects of *local* substructures which mostly affect a single lensed image from those of *distributed* substructures which affect multiple lensed images coherently. However, we generally expect these coherent perturbations to be somewhat degenerate with the smooth lens model. For instance, the dipolar ($p = 1$) component of the perturbation can generally be compensated by an appropriate shift to the source position, while the perturbation quadrupole ($p = 2$) could be reabsorbed by an appropriate external shear. We leave the study of these potential degeneracies to future work.

IV. ANALYTICAL MARGINALIZATION OVER DISTRIBUTED MASS SUBSTRUCTURES

In this section, we first describe how we transform from the linear lensing quantities to the actual gravitational lensing observables that can be compared with data. We then explain how we perform the analytical marginalization over the distributed mass substructure population using the characteristic function $Q_{\langle N_d \rangle}^{\text{dist}}(\mathbf{k}_L | \mathbf{q}, \mathbf{q}_{\text{sub}})$ computed in the previous section. Here, we assume that the effect of the local substructure population is perfectly known, which is equivalent to setting $Q_{\langle N_l \rangle}^{\text{local}}(\mathbf{k}_L | \mathbf{q}, \mathbf{q}_{\text{sub}}) = 1$. We also derive a general expression for the data likelihood in the presence of a distributed population of unresolved mass substructures. We first describe the general calculation and then specialize to the case where $Q_{\langle N_d \rangle}^{\text{dist}}(\mathbf{k}_L | \mathbf{q}, \mathbf{q}_{\text{sub}})$ is a multivariate Gaussian.

A. General case

The first step in to compute how the probability density function of the stochastic variables \mathbf{t} is related

to that of the stochastic variables $\Delta \mathbf{t}_L$ for which we have computed the characteristic function in the previous section. For definitiveness, we take $\Delta \mathbf{t}_L$ to contain all the stochastic deflections $\vec{\alpha}_{\text{sub}}^{(i)}$, $i \in N_{\text{img}}$, caused by distributed substructures as well as all the stochastic lensing potential shifts $\phi_{\text{sub}}^{(i)}$ between the i th image and the reference point. In the following, we take the reference point to be the position of the image that is leading the arrival time. We also take the stochastic vector of observables \mathbf{t} to contain the image positions $\mathbf{x}^{(i)}$ and the time delay $\Delta t^{(i)}$ between image i and the reference image caused by distributed substructures.

The probability density function for the perturbations $\Delta \mathbf{t}_L^{\text{dist}}$ to the linear lensing quantities caused by distributed substructures is simply

$$\Phi_{\langle N_d \rangle}(\Delta \mathbf{t}_L | \mathbf{q}, \mathbf{q}_{\text{sub}}) = \int \frac{d\mathbf{k}_L}{(2\pi)^l} Q_{\langle N_d \rangle}^{\text{dist}}(\mathbf{k}_L | \mathbf{q}, \mathbf{q}_{\text{sub}}) e^{-i\Delta \mathbf{t}_L \cdot \mathbf{k}_L}, \quad (43)$$

where it is understood the $\Delta \mathbf{t}_L$ stands for $\Delta \mathbf{t}_L^{\text{dist}}$ here. We then apply the transformation given by Eqs. (5) and (7) to compute the density function for the \mathbf{t} stochastic lensing observables

$$P(\mathbf{t} | \mathbf{q}, \mathbf{q}_{\text{sub}}) = \int d\mathbf{t}_L \int \frac{d\mathbf{k}_L}{(2\pi)^l} Q_{\langle N_d \rangle}^{\text{dist}}(\mathbf{k}_L | \mathbf{q}, \mathbf{q}_{\text{sub}}) \times e^{-i(\mathbf{t}_L - \bar{\mathbf{t}}_L) \cdot \mathbf{k}_L} \delta_D^k(\mathbf{t} - \mathbf{t}(\mathbf{t}_L)), \quad (44)$$

where k is the length of the \mathbf{t} vector and $\bar{\mathbf{t}}_L$ is the contribution to the linear lensing quantities from the smooth lens model [Eq. (6)]. Note that we generally have $l \neq k$ since, for instance, the two shear random variables are mapped to a single magnification perturbation. Nevertheless, as we are only considering potential and deflection perturbations, we have $l = k$ here. In practice, the relation $\mathbf{t}(\mathbf{t}_L)$ is nonlinear but can nonetheless be written down and inverted in a straightforward manner. However, since the distributed substructure population leads to small changes to the lensed image positions and their relative time delays, we can linearize this relation as

$$\mathbf{t}(\mathbf{t}_L) \approx \bar{\mathbf{t}} + \mathbf{A}^{-1}(\mathbf{t}_L - \bar{\mathbf{t}}_L), \quad (45)$$

where $\bar{\mathbf{t}}$ are the lensing observables in the absence of substructures, and where \mathbf{A} is a l by l matrix encoding the transformation between the linear lensing quantities and the actual observables. For instance, in the case where $\mathbf{t} = \{\mathbf{x}^{(i)}, \mathbf{x}^{(j)}, \mathbf{x}^{(k)}, \Delta t^{(j)}, \Delta t^{(k)}\}$ and $\Delta \mathbf{t}_L = \{\vec{\alpha}_{\text{sub}}^{(i)}, \vec{\alpha}_{\text{sub}}^{(j)}, \vec{\alpha}_{\text{sub}}^{(k)}, \phi_{\text{sub}}^{(j)}, \phi_{\text{sub}}^{(k)}\}$, and assuming that image i is the leading image, the inverse of the transformation matrix \mathbf{A} is given by

$$\mathbf{A} = \begin{pmatrix} \mu_s(\bar{\mathbf{x}}_i)^{-1} & 0 & 0 & 0 & 0 \\ 0 & \mu_s(\bar{\mathbf{x}}_j)^{-1} & 0 & 0 & 0 \\ 0 & 0 & \mu_s(\bar{\mathbf{x}}_k)^{-1} & 0 & 0 \\ 0 & 0 & 0 & -t_0^{-1} & 0 \\ 0 & 0 & 0 & 0 & -t_0^{-1} \end{pmatrix}, \quad (46)$$

where $\mu_s(\bar{\mathbf{x}}_i)$ stands for the 2×2 magnification tensor of the smooth lens component evaluated at the unperturbed image position $\bar{\mathbf{x}}_i$, and t_0 is the time constant of the lens which is given by

$$t_0 = \frac{1 + z_{\text{lens}}}{c} \frac{D_l D_s}{D_{ls}}, \quad (47)$$

where z_{lens} is the redshift of the lens, c is the speed of light, and D_l , D_s , and D_{ls} are the angular diameter distances to the lens, to the source, and from the lens to the source, respectively. As it should be apparent from Eq. (46), the matrix $\mathbf{A} \equiv \mathbf{A}(\mathbf{q}_{\text{gal}}, \mathbf{q}_{\text{env}}, \mathbf{q}_{\text{cos}})$ depends only the smooth mass component of the lens, its environment, and the cosmological model. Substituting Eq. (45) into Eq. (44), we can then perform the \mathbf{t}_L integration

$$P(\mathbf{t}|\mathbf{q}, \mathbf{q}_{\text{sub}}) = |\mathbf{A}| \int \frac{d\mathbf{k}_L}{(2\pi)^l} Q_{\langle N_d \rangle}^{\text{dist}}(\mathbf{k}_L|\mathbf{q}, \mathbf{q}_{\text{sub}}) e^{-i(\mathbf{t}-\bar{\mathbf{t}})^T \mathbf{A}^T \mathbf{k}_L}, \quad (48)$$

where $|\mathbf{A}|$ stands for the determinant of the matrix \mathbf{A} . We can finally substitute the above into Eq. (2) to compute the data likelihood in the presence of a distributed population of substructures

$$\begin{aligned} \mathcal{L}(\mathbf{d}|\mathbf{q}, \mathbf{q}_{\text{sub}}) &= \frac{|\mathbf{A}|}{\sqrt{(2\pi)^l |\mathbf{C}_d|}} \int d\mathbf{t} \int \frac{d\mathbf{k}_L}{(2\pi)^l} Q_{\langle N_d \rangle}^{\text{dist}}(\mathbf{k}_L|\mathbf{q}, \mathbf{q}_{\text{sub}}) \\ &\quad \times e^{-i(\mathbf{t}-\bar{\mathbf{t}})^T \mathbf{A}^T \mathbf{k}_L} e^{-\frac{1}{2}(\mathbf{t}-\mathbf{d})^T \mathbf{C}_d^{-1}(\mathbf{t}-\mathbf{d})} \\ &= |\mathbf{A}| \int \frac{d\mathbf{k}_L}{(2\pi)^l} Q_{\langle N_d \rangle}^{\text{dist}}(\mathbf{k}_L|\mathbf{q}, \mathbf{q}_{\text{sub}}) e^{-\frac{1}{2}\mathbf{k}_L^T \mathbf{A} \mathbf{C}_d \mathbf{A}^T \mathbf{k}_L} \\ &\quad \times e^{-i(\mathbf{d}-\bar{\mathbf{t}})^T \mathbf{A}^T \mathbf{k}_L}, \end{aligned} \quad (49)$$

where we have assumed a Gaussian likelihood, $\mathcal{L}(\mathbf{d}|\mathbf{t}) \propto \exp[-\frac{1}{2}(\mathbf{t}-\mathbf{d})^T \mathbf{C}_d^{-1}(\mathbf{t}-\mathbf{d})]$, where \mathbf{C}_d is the data covariance matrix. Thus, the data likelihood is given by the Fourier transform of the product of $Q_{\langle N_d \rangle}^{\text{dist}}(\mathbf{k}_L|\mathbf{q}, \mathbf{q}_{\text{sub}})$ with the Fourier conjugate distribution of $\mathcal{L}(\mathbf{d}|\mathbf{t})$ evaluated at the residuals $\mathbf{d}-\bar{\mathbf{t}}$ between the data and the predictions from the smooth lens model. This makes sense: The \mathbf{k}_L modes contributing most to the integral are those that can explain the residuals

between the actual data and the smooth lens model, and the characteristic function $Q_{\langle N_d \rangle}^{\text{dist}}$ encodes whether these modes are likely to contribute to the residuals given the input substructure properties.

B. Gaussian case

We now specialize to the case where $Q_{\langle N_d \rangle}^{\text{dist}}(\mathbf{k}_L|\mathbf{q}, \mathbf{q}_{\text{sub}})$ is well approximated by a multivariate Gaussian. Starting from Eq. (49), we have

$$\begin{aligned} \mathcal{L}(\mathbf{d}|\mathbf{q}, \mathbf{q}_{\text{sub}}) &= |\mathbf{A}| \int \frac{d\mathbf{k}_L}{(2\pi)^l} e^{-\frac{1}{2}\mathbf{k}_L^T \mathbf{C}_{\text{sub}} \mathbf{k}_L} e^{-\frac{1}{2}\mathbf{k}_L^T \mathbf{A} \mathbf{C}_d \mathbf{A}^T \mathbf{k}_L} \\ &\quad \times e^{-i(\mathbf{d}-\bar{\mathbf{t}})^T \mathbf{A}^T \mathbf{k}_L} \\ &= \frac{|\mathbf{A}|}{\sqrt{(2\pi)^l |\mathbf{C}_{\text{sub}} + \mathbf{A} \mathbf{C}_d \mathbf{A}^T|}} \\ &\quad \times e^{-\frac{1}{2}(\mathbf{d}-\bar{\mathbf{t}})^T \mathbf{A}^T (\mathbf{C}_{\text{sub}} + \mathbf{A} \mathbf{C}_d \mathbf{A}^T)^{-1} \mathbf{A} (\mathbf{d}-\bar{\mathbf{t}})}, \end{aligned} \quad (50)$$

thus leading to a Gaussian data likelihood with total inverse covariance matrix given by

$$\mathbf{C}_{\text{tot}}^{-1} = \mathbf{A}^T (\mathbf{C}_{\text{sub}} + \mathbf{A} \mathbf{C}_d \mathbf{A}^T)^{-1} \mathbf{A}. \quad (51)$$

This analysis shows that the effect of distributed unresolved mass substructures can be thought of as an additional source of noise that directly contributes to modeling uncertainty. This extra contribution to the net covariance matrix entering the likelihood describes the inherent mass modeling uncertainties due to the lumpiness of massive galaxies acting as strong gravitational lenses. Whether the inherent mass modeling uncertainties caused by mass substructures are relevant or not in the above likelihood depend on the observational precision of a given data set. Conversely, in the event that precise time delay and astrometric observations are available, they can be used with the above likelihood to constrain the physical quantities entering the covariance matrix \mathbf{C}_{sub} .

V. DISCUSSION

The analytical approximation developed in Sec. III B allows us to not only marginalize over the masses and positions of distributed substructures but also to understand which of their physical properties are most relevant to gravitational lensing observables. For the physically relevant parameter space, the effect of distributed mass substructures on the lensing potential and its first derivative is approximately Gaussian, which implies that most of the relevant physics is encoded in the covariance matrix \mathbf{C}_{sub} . Using Eqs. (21) and (C1), this covariance matrix can be decomposed as follow

$$\mathbf{C}_{\text{sub}}^{ij} = \langle N_d \rangle \langle m^2 \rangle \langle \mathcal{O}_L^i \mathcal{O}_L^j \rangle = \langle \kappa_{\text{sub}}(r_{\text{ref}}) \rangle \frac{\langle m^2 \rangle}{\langle m \rangle} \frac{\langle \mathcal{O}_L^i \mathcal{O}_L^j \rangle}{\mathcal{P}_r(r_{\text{ref}})}, \quad (52)$$

where r_{ref} is a reference radius where the amplitude of the convergence in mass substructures is set (taken to be the Einstein radius of the main lens in earlier sections of this paper), and where

$$\langle \mathcal{O}_L^i \mathcal{O}_L^j \rangle \equiv \int d^2r \mathcal{P}_r(r, \theta) \mathcal{O}_L^i \mathcal{O}_L^j. \quad (53)$$

In going from the first to the second equality in Eq. (52), we use Eqs. (40) and (41) to express $\langle N_d \rangle$ in terms of κ_{sub} . We give in Appendix D useful expressions for the different entries of \mathbf{C}_{sub} . The expected number of distributed substructures is given by Eq. (40), which for $\beta < -1$ and $M_{\text{low}} \ll M_{\text{high}}$ is approximately given by

$$\langle N_d \rangle \sim -\frac{a_0 M_{\text{low}}^{\beta+1}}{M_0^\beta (\beta+1)} (1 - \mathcal{P}_r(< R_{\text{min}})). \quad (54)$$

The second moment of the mass function $\langle m^2 \rangle$ is easily computed from Eq. (20),

$$\langle m^2 \rangle = \frac{1}{\pi^2 \Sigma_{\text{crit}}^2} \frac{\beta+1}{\beta+3} \frac{M_{\text{high}}^{\beta+3} - M_{\text{low}}^{\beta+3}}{M_{\text{high}}^{\beta+1} - M_{\text{low}}^{\beta+1}} \sim -\frac{1}{\pi^2 \Sigma_{\text{crit}}^2} \frac{\beta+1}{\beta+3} \frac{M_{\text{high}}^{\beta+3}}{M_{\text{low}}^{\beta+1}} \quad (55)$$

for $M_{\text{low}} \ll M_{\text{high}}$,

where the last approximation is valid for when $-3 < \beta < -1$. The second moment of the mass function thus has a rather strong dependence on the minimal subhalo mass. Now, if we look at the product $\langle N_d \rangle \langle m^2 \rangle$, we immediately see that the leading dependence on M_{low} cancels out for the physically relevant case $-3 < \beta < -1$ and $M_{\text{low}} \ll M_{\text{high}}$

$$\begin{aligned} \langle N_d \rangle \langle m^2 \rangle &= \frac{1}{\pi^2 \Sigma_{\text{crit}}^2} \frac{a_0}{\beta+3} \frac{M_{\text{high}}^{\beta+3} - M_{\text{low}}^{\beta+3}}{M_0^\beta} (1 - \mathcal{P}_r(< R_{\text{min}})) \\ &\sim \frac{1}{\pi^2 \Sigma_{\text{crit}}^2} \frac{a_0}{\beta+3} \frac{M_{\text{high}}^{\beta+3}}{M_0^\beta} (1 - \mathcal{P}_r(< R_{\text{min}})). \end{aligned} \quad (56)$$

This shows that the scaling of the substructure covariance matrix depends mostly on the normalization of the mass function a_0 and on the largest subhalo mass M_{high} . We illustrate this behavior in Fig. 7 for different values of the mass function slope β . In each case, we choose the mass function normalization such that all curves asymptote to the same value as $M_{\text{low}}/M_{\text{high}} \rightarrow 0$. In the regime where Gaussianity holds ($M_{\text{low}}/M_{\text{high}} \lesssim 0.1$), we

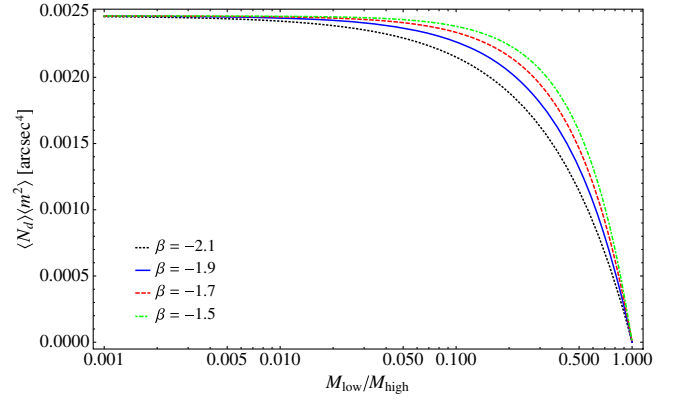


FIG. 7. Dependence of the product $\langle N_d \rangle \langle m^2 \rangle$ on the $M_{\text{low}}/M_{\text{high}}$ for four values of the slope of the mass function. In each case, we choose the normalization of the subhalo mass function a_0 such that all curves asymptote to the same value as $M_{\text{low}}/M_{\text{high}} \rightarrow 0$. Our choice of normalization corresponds to $a_0 = 3.8 \times 10^{-10} M_\odot^{-1}$ at $M_0 = M_{\text{high}} = 10^{10} M_\odot$ when $\beta = -1.9$.

observe that $\langle N_d \rangle \langle m^2 \rangle$ is roughly constant as $M_{\text{low}}/M_{\text{high}}$ is varied. Measurement of this constant could provide a consistency test for standard cold dark matter theory. Since there are strong degeneracies between the different parameters in Eq. (56), the extraction of individual mass function parameters would require strong priors from either simulations or complementary observations.

In addition to its dependence on the subhalo mass function, the covariance matrix \mathbf{C}_{sub} also encodes important information about the spatial distribution of distributed mass substructures. Since each entry of the covariance matrix depends on the spatial distribution through a different kernel [Eq. (53)], it is reasonable to believe that lensing observables will provide good sensitivity to the spatial distribution of substructures. A detailed analysis of the sensitivity of different lensing observations to unresolved substructures will be carried in an upcoming work. Looking ahead, we expect that combining magnification information (mostly sensitive to local substructures), with astrometric fluctuations (sensitive to both local and distributed substructures) and time delay perturbations (sensitive to distributed substructures with some local sensitivity) will lead to a comprehensive picture of the satellite populations of distant lens galaxies.

VI. CONCLUSION

In this paper, we have computed from first principles the probability distribution of lensing potentials and deflections in the presence of an unresolved population of mass substructures that are located

beyond the strong lensing region. We have determined that for a realistic substructure population, the distribution of lensing potential and deflection perturbations is close to a multivariate Gaussian. We have computed the leading order deviations from Gaussianity and used them to determine when the probability distribution ceases to be well approximated by Gaussian statistics. We have shown in Sec. IV how our technique can be used to efficiently marginalize over the properties of distributed mass substructures without having to perform costly numerical simulations of mass substructure populations.

For simplicity, we have treated distributed substructures as independent point masses, which we believe is an excellent approximation for subhalos far away from lensed images. We note that our approach can straightforwardly be generalized to clustered and extended substructures if we compute the theory covariance matrix as

$$C_{\text{sub}}^{ij} \propto \int d^2r \int d^2r' K_i(\mathbf{r}) K_j(\mathbf{r}') \langle \kappa_{\text{sub}}(\mathbf{r}) \kappa_{\text{sub}}(\mathbf{r}') \rangle, \quad (57)$$

where the kernels K_i and K_j depend on which linear lensing quantities are being used. We note that in the point-mass limit the above expression reduces to Eq. (52). Therefore, we can see that, in the general case, we are really probing the ensemble-averaged two-point correlation function of the distributed substructure convergence field. This two-point function, which is in general neither homogeneous nor isotropic, can be directly measured in N -body or semi-analytic simulations, hence providing a way to assess the importance of subhalo clustering (the two-halo term) and to test the accuracy of the point-mass approximation. We leave such tests to future work.

In the present manuscript, we have focused our attention on time delay and astrometry perturbations since these are the lensing observables that are most sensitive to distributed mass substructures. Expanding our analysis to include mass substructures near lensed images would allow the incorporation of magnification information into our framework. Together, the relative flux measurements, positions, and time delays between lensed images have power to constrain both the local and distributed substructure populations of a lens galaxy, given appropriate levels of measurement precision. Quantifying these precision levels in detail will vary from system to system and will be the subject of future work. As an example, in the case of time delays, the fluctuations caused by distributed substructures are demonstrated to be $\sqrt{\Delta t^2} < 1$ day [42], suggesting time delay precision levels on the order of hours.

One of the main advantages of having an analytical framework to handle mass substructures is that it allows efficient exploration of degeneracies between substructure effects (\mathbf{q}_{sub}), on the one hand, and the macrolens (\mathbf{q}_{gal}), its environment (\mathbf{q}_{env}) and possible line-of-sight structures (\mathbf{q}_{los}) on the other. Exploring and marginalizing over these degeneracies is important in assessing the sensitivity of current and future data to the detection of a population of nonluminous mass substructures in the outskirts of distant galaxies. Such a detection would confirm a key prediction, or offer a quantitative challenge, of our current paradigm for structure formation based on the CDM scenario. The synthesis of all lensing observables which are sensitive to different combinations of local and distant substructures, measured with sufficient precision, have the potential to produce a complete picture of the substructure mass function. The stochastic millilensing framework developed here is a necessary step toward this goal.

ACKNOWLEDGMENTS

We thank Geoffrey Bryden, James Bullock, Curt Cutler, Olivier Doré, David Hogg, Jeffrey Jewel, James Taylor, and Michele Vallisneri for useful conversations. The work of F.-Y. C.-R. was performed in part at the California Institute of Technology for the Keck Institute for Space Studies, which is funded by the W. M. Keck Foundation. F.-Y. C.-R. thanks the Aspen Center for Physics, where some of this work was performed. The Aspen Center for Physics is supported by the National Science Foundation under Grant No. 1066293. Part of the research described in this paper was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration (NASA). L. A. M. gratefully acknowledges support by the NASA ATFP program through Award No. 399131.02.02.02.98. C. R. K. acknowledges support from the National Science Foundation under Grant No. AST-0747311. The research of K. S. is supported in part by a Natural Science and Engineering Research Council (NSERC) of Canada Discovery Grant. D. A. G. acknowledges the support of the NASA Undergraduate Internship and Student Internship programs.

APPENDIX A: SUBHALO SCALE AND TRUNCATION RADII

1. Scale radius

In this appendix, we derive a relation between the total mass M_{sub} of a subhalo and its scale radius r_s . Our starting point is the relation between the maximum circular velocity inside a subhalo v_{max} and the radius r_{max} at which this velocity occurs [124]

$$\left(\frac{r_{\max}}{1 \text{ kpc}}\right) = (0.72 \pm 0.25) \left(\frac{v_{\max}}{10 \text{ km s}^{-1}}\right)^{1.47}. \quad (\text{A1})$$

By the virial theorem, we also have

$$\frac{GM_{\text{sub}}(<r_{\max})}{r_{\max}} = v_{\max}^2, \quad (\text{A2})$$

where G is the gravitational constant and $M_{\text{sub}}(<r_{\max})$ is the subhalo mass enclosed within r_{\max} , which for a smoothly truncated NFW density profile given in Eq. (13) is

$$M_{\text{sub}}(<r_{\max}) = M_{\text{sub}} \frac{4(1+x_{\max})\tau \arctan(x_{\max}/\tau) - 2x_{\max}(1+\tau^2) + (1+x_{\max})(\tau^2-1) \ln\left[\frac{\tau^2(1+x_{\max})^2}{(x_{\max}^2+\tau^2)}\right]}{2(1+x_{\max})((\tau^2-1)\ln\tau + \pi\tau - (\tau^2+1))}, \quad (\text{A3})$$

where M_{sub} is the total mass of the subhalo given in Eq. (14), $\tau \equiv r_t/r_s$ (where r_t is the tidal radius) and $x_{\max} \equiv r_{\max}/r_s$. We substitute the above expression into Eq. (A2) and maximize the left-hand side to find the radius at which the maximum circular velocity occurs. The resulting equation is not solvable analytically, but for realistic values of τ , we obtain

$$x_{\max} \approx 2.1626 \left(1 - \frac{1}{1+\tau^2}\right)^2 \quad \text{for } \tau \gtrsim 7. \quad (\text{A4})$$

We then substitute the above into Eqs. (A3) and (A2) and use Eq. (A1) to eliminate v_{\max} from Eq. (A2) in order to obtain a relation between the scale radius r_s and the total mass of the subhalo M_{sub}

$$\left(\frac{r_s}{1 \text{ kpc}}\right) = 1.0 \pm 0.3 \left(\frac{M_{\text{sub}}}{10^9 M_{\odot}}\right)^{0.735}, \quad (\text{A5})$$

where we have neglected a weak dependence on τ since it leads to changes that are smaller than the scatter about the mean.

2. Tidal truncation radius

We use the result of Refs. [125,126] for the tidal truncation radius r_t of a subhalo of mass M_{sub} located at a distance r_{3D} from the main halo center,

$$r_t = \left(\frac{M_{\text{sub}}}{[2 - d \ln M_{\text{main}}/d \ln r_{3D}] M_{\text{main}}(<r_{3D})}\right)^{1/3} r_{3D}, \quad (\text{A6})$$

where M_{main} is the mass of the main lens halo and where both r_t and r_{3D} are radii in three-dimensional space (not projected). Since the truncation radius is most relevant to the stochastic lensing signal for subhalos lying close or within the Einstein radius of the lens, it is sufficient to specify the mass distribution in the vicinity of the lensed images. In this work, we focus on power-law mass models which have been

shown to provide good fits to many gravitational lenses. The projected mass density divided by the critical density for lensing for these models is given by [132]

$$\kappa_{\text{main}}(r) = \frac{1}{2} \left(\frac{b}{r}\right)^{2-\alpha_{\text{main}}}, \quad (\alpha_{\text{main}} \neq 2), \quad (\text{A7})$$

where b is a length scale closely related to the Einstein radius of the main lens and α_{main} is the power-law index of the density profile. We can deproject this relation to obtain the 3-d mass profile of the main lens

$$\rho_{\text{main}}(r_{3D}) = \frac{1}{2\sqrt{\pi}b} \Sigma_{\text{crit}} \frac{\Gamma(\frac{3-\alpha_{\text{main}}}{2})}{\Gamma(\frac{2-\alpha_{\text{main}}}{2})} \left(\frac{b}{r_{3D}}\right)^{3-\alpha_{\text{main}}}, \quad (\text{A8})$$

where $\Gamma(x)$ is the gamma function. This relation is easily integrated to obtain $M_{\text{main}}(<r_{3D})$, which leads a tidal truncation radius given by

$$r_t = \left(\frac{\alpha_{\text{main}}}{2-\alpha_{\text{main}}} \frac{\Gamma(\frac{2-\alpha_{\text{main}}}{2})}{\Gamma(\frac{3-\alpha_{\text{main}}}{2})} \frac{M_{\text{sub}}}{2\sqrt{\pi}\Sigma_{\text{crit}}b^2}\right)^{1/3} \left(\frac{b}{r_{3D}}\right)^{\frac{\alpha_{\text{main}}}{3}} r_{3D}, \quad (\alpha_{\text{main}} \neq 2). \quad (\text{A9})$$

We are generally interested in quasi-isothermal inner density profiles ($\alpha_{\text{main}} \sim 1$) for the main lens halo, which implies that we generally have

$$r_t \propto M_{\text{sub}}^{1/3} r_{3D}^{2/3} \quad (\text{A10})$$

for a subhalo of mass M_{sub} located at a distance r_{3D} from the center of the main lens halo.

APPENDIX B: SPATIAL MOMENTS OF THE SUBSTRUCTURE DISTRIBUTION

In this appendix, we simplify the structure of the spatial integral appearing in Eq. (33) using a multinomial expansion. Keeping only the first N_{mult} multipoles in Eq. (29), the spatial integral takes the form

$$\begin{aligned}
\int_{\mathcal{H}_d} d^2r \mathcal{P}_r(r, \theta; \mathbf{q}_{\text{sub}}) (\mathbf{k}_L \cdot \vec{\mathcal{O}}_L)^n &= \int_{\mathcal{H}_d} d^2r \mathcal{P}_r(r, \theta; \mathbf{q}_{\text{sub}}) \left(- \sum_{p=1}^{N_{\text{mult}}} \frac{1}{r^p} [\mathbf{k}_L \cdot \vec{A}_p \cos(p\theta) + \mathbf{k}_L \cdot \vec{B}_p \sin(p\theta)] \right)^n \\
&= (-1)^n \int_{\mathcal{H}_d} d^2r \mathcal{P}_r(r, \theta; \mathbf{q}_{\text{sub}}) \left(\sum_{\sum p_i = n} \binom{n}{p_1, p_2, \dots, p_{2N_{\text{mult}}}} \right. \\
&\quad \times \prod_{t=1}^{N_{\text{mult}}} \left[\frac{1}{r^t} \mathbf{k}_L \cdot \vec{A}_t \cos(t\theta) \right]^{p_t} \prod_{s=1}^{N_{\text{mult}}} \left[\frac{1}{r^s} \mathbf{k}_L \cdot \vec{B}_s \sin(s\theta) \right]^{p_{N_{\text{mult}}+s}} \Bigg) \\
&= (-1)^n \sum_{\|\mathbf{p}\|=n} \binom{n}{\mathbf{p}} K_{\mathbf{p}} \prod_{t=1}^{N_{\text{mult}}} (\mathbf{k}_L \cdot \vec{A}_t)^{p_t} \prod_{s=1}^{N_{\text{mult}}} (\mathbf{k}_L \cdot \vec{B}_s)^{p_{N_{\text{mult}}+s}}, \tag{B1}
\end{aligned}$$

where $\mathbf{p} = \{p_1, p_2, \dots, p_{2N_{\text{mult}}}\}$ is a multi-index with $\|\mathbf{p}\| = \sum_{j=1}^{2N_{\text{mult}}} p_j$, and where the kernel $K_{\mathbf{p}}$ is given by

$$K_{\mathbf{p}} = \int_{\mathcal{H}_d} d^2r \mathcal{P}_r(r, \theta; \mathbf{q}_{\text{sub}}) \left(\frac{1}{r} \right)^{\sum_{j=1}^{2N_{\text{mult}}} p_j} \prod_{t=1}^{N_{\text{mult}}} \cos(t\theta)^{p_t} \prod_{s=1}^{N_{\text{mult}}} \sin(s\theta)^{p_{N_{\text{mult}}+s}}, \tag{B2}$$

and where

$$\binom{n}{\mathbf{p}} = \frac{n!}{p_1! p_2! \dots p_{2N_{\text{mult}}}!} \tag{B3}$$

is the multinomial coefficient.

APPENDIX C: EDGEWORTH EXPANSION OF THE CHARACTERISTIC FUNCTION

In this appendix, we perform the Edgeworth expansion of the characteristic function $\mathcal{Q}_{\langle N_d \rangle}^{\text{dist}}$ in powers of $\langle N_d \rangle^{-1/2}$. We start by writing the Cholesky decomposition of the covariance matrix \mathbf{C}_{sub} as

$$\mathbf{C}_{\text{sub}} = \langle N_d \rangle \langle m^2 \rangle \Lambda \Lambda^T, \tag{C1}$$

which is always possible since \mathbf{C}_{sub} is a symmetric positive-definite matrix, and where we have pulled out the overall scaling with the average number of mass substructures and the second moment of the mass function. We can then define a new normalized Fourier conjugate variable $\tilde{\mathbf{k}}_L$ as

$$\tilde{\mathbf{k}}_L = \sqrt{\langle N_d \rangle \langle m^2 \rangle} \Lambda^T \mathbf{k}_L \tag{C2}$$

and express the characteristic function as a function of it. Since the characteristic function $\tilde{\mathcal{Q}}_{\langle N_d \rangle}^{\text{dist}}(\tilde{\mathbf{k}}_L)$ of the normalized variable $\tilde{\mathbf{k}}_L$ is related to that given in Eq. (36) by

$$\tilde{\mathcal{Q}}_{\langle N_d \rangle}^{\text{dist}}(\tilde{\mathbf{k}}_L) = \frac{1}{\sqrt{\langle N_d \rangle \langle m^2 \rangle} |\Lambda|} \mathcal{Q}_{\langle N_d \rangle}^{\text{dist}} \left(\frac{(\Lambda^T)^{-1} \tilde{\mathbf{k}}_L}{\sqrt{\langle N_d \rangle \langle m^2 \rangle}} \right), \tag{C3}$$

we obtain

$$\begin{aligned}
\tilde{\mathcal{Q}}_{\langle N_d \rangle}^{\text{dist}}(\tilde{\mathbf{k}}_L) &= \frac{e^{-\frac{1}{2} \tilde{\mathbf{k}}_L \cdot \tilde{\mathbf{k}}_L}}{\sqrt{\langle N_d \rangle \langle m^2 \rangle} |\Lambda|} \exp \left[\sum_{n=3}^{\infty} \frac{(-1)^n \langle m^n \rangle}{n! \langle N_d \rangle^{n/2-1} \langle m^2 \rangle^{n/2}} \right. \\
&\quad \times \left. \left(\sum_{\|\mathbf{p}\|=n} \binom{n}{\mathbf{p}} K_{\mathbf{p}} \prod_{t=1}^{N_{\text{mult}}} (i \tilde{\mathbf{k}}_L^T \Lambda^{-1} \vec{A}_t)^{p_t} \prod_{s=1}^{N_{\text{mult}}} (i \tilde{\mathbf{k}}_L^T \Lambda^{-1} \vec{B}_s)^{p_{N_{\text{mult}}+s}} \right) \right], \tag{C4}
\end{aligned}$$

where we have taken the mean vector $\mathbf{u} \equiv \langle \Delta \mathbf{t}_L \rangle$ to vanish, but the above result can straightforwardly be generalized to a nonzero mean values of linear lensing quantities. To make the notation more compact, we define

$$\langle (i \tilde{\mathbf{k}}_L \cdot \vec{\mathcal{O}}_L)^n \rangle \equiv \sum_{\|\mathbf{p}\|=n} \binom{n}{\mathbf{p}} K_{\mathbf{p}} \prod_{t=1}^{N_{\text{mult}}} (i \tilde{\mathbf{k}}_L^T \Lambda^{-1} \vec{A}_t)^{p_t} \prod_{s=1}^{N_{\text{mult}}} (i \tilde{\mathbf{k}}_L^T \Lambda^{-1} \vec{B}_s)^{p_{N_{\text{mult}}+s}}. \tag{C5}$$

We then expand the exponential in Eq. (C4) to obtain the proper Edgeworth expansion of $\tilde{\mathcal{Q}}_{\langle N_d \rangle}^{\text{dist}}(\tilde{\mathbf{k}}_L)$

$$\begin{aligned} \tilde{Q}_{\langle N_d \rangle}^{\text{distant}}(\tilde{\mathbf{k}}_L) = & \frac{e^{-\frac{1}{2}\tilde{\mathbf{k}}_L \cdot \tilde{\mathbf{k}}_L}}{\sqrt{\langle N_d \rangle \langle m^2 \rangle} |\Lambda|} \left(1 - \frac{1}{\langle N_d \rangle^{1/2}} \left[\frac{\langle m^3 \rangle \langle (i\tilde{\mathbf{k}}_L \cdot \vec{\mathcal{O}}_L)^3 \rangle}{3! \langle m^2 \rangle^{3/2}} \right] + \frac{1}{\langle N_d \rangle} \left[\frac{\langle m^4 \rangle \langle (i\tilde{\mathbf{k}}_L \cdot \vec{\mathcal{O}}_L)^4 \rangle}{4! \langle m^2 \rangle^2} + \frac{(\langle m^3 \rangle \langle (i\tilde{\mathbf{k}}_L \cdot \vec{\mathcal{O}}_L)^3 \rangle)^2}{72 \langle m^2 \rangle^3} \right] \right. \\ & - \frac{1}{\langle N_d \rangle^{3/2}} \left[\frac{\langle m^5 \rangle \langle (i\tilde{\mathbf{k}}_L \cdot \vec{\mathcal{O}}_L)^5 \rangle}{5! \langle m^2 \rangle^{5/2}} + \frac{\langle m^3 \rangle \langle (i\tilde{\mathbf{k}}_L \cdot \vec{\mathcal{O}}_L)^3 \rangle \langle m^4 \rangle \langle (i\tilde{\mathbf{k}}_L \cdot \vec{\mathcal{O}}_L)^4 \rangle}{144 \langle m^2 \rangle^{7/2}} + \frac{(\langle m^3 \rangle \langle (i\tilde{\mathbf{k}}_L \cdot \vec{\mathcal{O}}_L)^3 \rangle)^3}{1296 \langle m^2 \rangle^{9/2}} \right] \\ & \left. + O(\langle N_d \rangle^{-2}) \right). \end{aligned} \quad (\text{C6})$$

APPENDIX D: COVARIANCE MATRIX FOR LINEAR LENSING QUANTITIES

1. General expressions in the presence of circular symmetry

In general, the covariance matrix for linear lensing quantities is given by

$$\mathbf{C}_{\text{sub}}^{ij} = \langle N_d \rangle \langle m^2 \rangle \int d^2r \mathcal{P}_r(r, \theta) \left[\sum_{p=1}^{\infty} \frac{1}{r^p} [A_p^{(i)} \cos(p\theta) + B_p^{(i)} \sin(p\theta)] \right] \left[\sum_{t=1}^{\infty} \frac{1}{r^t} [A_t^{(j)} \cos(t\theta) + B_t^{(j)} \sin(t\theta)] \right], \quad (\text{D1})$$

where $A_p^{(i)}$ denotes the i th component of the vector \vec{A}_p . In the case of a circular halo with $\mathcal{P}_r(r, \theta) \equiv \mathcal{P}_r(r)$, the above expression dramatically simplifies. In the following, we provide convenient expressions for the different entries of the covariance matrix for the linear lensing quantities. We take the position of image i to be $\mathbf{x}_i = (r_i \cos \theta_i, r_i \sin \theta_i)$, that of image j to be $\mathbf{x}_j = (r_j \cos \theta_j, r_j \sin \theta_j)$, and the reference point for the projected lensing potential is $\mathbf{x}_{\text{ref}} = (r_{\text{ref}} \cos \theta_{\text{ref}}, r_{\text{ref}} \sin \theta_{\text{ref}})$. We use the notation $\phi_{\text{sub}}^{(i)}$ to denote the difference in projected potential between image i and the reference point, that is, $\phi_{\text{sub}}^{(i)} \equiv \phi_{\text{sub}}(\mathbf{x}_i) - \phi_{\text{sub}}(\mathbf{x}_{\text{ref}})$. The covariances for the deflections are given by

$$\langle \alpha_{\text{sub},x}^{(i)} \alpha_{\text{sub},x}^{(j)} \rangle = \langle \alpha_{\text{sub},y}^{(i)} \alpha_{\text{sub},y}^{(j)} \rangle = \langle N_d \rangle \langle m^2 \rangle \sum_{p=0}^{\infty} r_i^p r_j^p \mathcal{K}[2(p+1)] \cos\{p(\theta_i - \theta_j)\}, \quad (\text{D2})$$

$$\langle \alpha_{\text{sub},x}^{(i)} \alpha_{\text{sub},y}^{(j)} \rangle = \langle N_d \rangle \langle m^2 \rangle \sum_{p=1}^{\infty} r_i^p r_j^p \mathcal{K}[2(p+1)] \sin\{p(\theta_i - \theta_j)\}, \quad (\text{D3})$$

where the kernel $\mathcal{K}[n]$ is given by

$$\mathcal{K}[n] \equiv \pi \int_{R_{\text{min}}}^{R_{\text{max}}} dr r \mathcal{P}_r(r) \frac{1}{r^n}, \quad (\text{D4})$$

where the leading factor of π comes from the angular integration over θ . We give explicit expressions for $\mathcal{K}[n]$ in the next subsection for two choices of spatial distributions. The cross terms between deflections and lensing potential are

$$\langle \alpha_{\text{sub},x}^{(i)} \phi_{\text{sub}}^{(j)} \rangle = \langle N_d \rangle \langle m^2 \rangle \sum_{p=1}^{\infty} \frac{r_i^{p-1} \mathcal{K}[2p]}{p} (r_j^p \cos[(p-1)\theta_i - p\theta_j] - r_{\text{ref}}^p \cos[(p-1)\theta_i - p\theta_{\text{ref}}]), \quad (\text{D5})$$

$$\langle \alpha_{\text{sub},y}^{(i)} \phi_{\text{sub}}^{(j)} \rangle = \langle N_d \rangle \langle m^2 \rangle \sum_{p=1}^{\infty} \frac{r_i^{p-1} \mathcal{K}[2p]}{p} (-r_j^p \sin[(p-1)\theta_i - p\theta_j] + r_{\text{ref}}^p \sin[(p-1)\theta_i - p\theta_{\text{ref}}]). \quad (\text{D6})$$

Finally, the covariance between projected lensing potential is given by

$$\langle \phi_{\text{sub}}^{(i)} \phi_{\text{sub}}^{(j)} \rangle = \langle N_d \rangle \langle m^2 \rangle \sum_{p=1}^{\infty} \frac{\mathcal{K}[2p]}{p^2} (r_i^p r_j^p \cos[p(\theta_i - \theta_j)] + r_{\text{ref}}^p (r_{\text{ref}}^p - r_i^p \cos[p(\theta_i - \theta_{\text{ref}})] - r_j^p \cos[p(\theta_j - \theta_{\text{ref}})])). \quad (\text{D7})$$

2. \mathcal{K} kernel for different choices of spatial distributions

We now provide explicit expressions for the spatial kernel given in Eq. (D4) above.

a. Power-law spatial distribution

We consider the following power-law spatial distribution

$$\mathcal{P}_r(r) = \frac{\eta r^{\eta-2}}{2\pi(R_{\max}^\eta - R_{\min}^\eta)}, \quad (0 < \eta \leq 2), \quad (\text{D8})$$

where the case $\eta = 1$ corresponds to an isothermal profile, while $\eta = 2$ reduces to the case of a uniform spatial distribution. For this distribution, the kernel takes the form

$$\mathcal{K}[n] = \begin{cases} \frac{\eta(R_{\min}^{\eta-n} - R_{\max}^{\eta-n})}{2(n-\eta)(R_{\max}^\eta - R_{\min}^\eta)} & \text{if } n \neq \eta, \\ \frac{\eta}{2(R_{\max}^\eta - R_{\min}^\eta)} \ln\left(\frac{R_{\max}}{R_{\min}}\right) & \text{if } n = \eta. \end{cases} \quad (\text{D9})$$

b. Cored spatial distribution

In this case, the spatial distribution of substructure is given by

$$\mathcal{P}_r(r) = \left(\frac{1}{2\pi r_c^2} \frac{1}{W(R_{\max}/r_c) - W(R_{\min}/r_c)} \right) \times \frac{1}{(1 + (r/r_c))^2},$$

$$\text{where } W(x) = \frac{1}{1+x} + \ln(1+x), \quad (\text{D10})$$

and where r_c is the core radius. The kernel is then given by

$$\begin{aligned} \mathcal{K}[n] = & \frac{1}{2} \frac{1}{W(R_{\max}/r_c) - W(R_{\min}/r_c)} \\ & \times \left[\frac{R_{\min}^{1-n}}{r_c + R_{\min}} - \frac{R_{\max}^{1-n}}{r_c + R_{\max}} \right. \\ & + \frac{(n-1)}{n} \left[\frac{1}{R_{\max}^n} {}_2F_1\left(1, n; 1+n; -\frac{r_c}{R_{\max}}\right) \right. \\ & \left. \left. - \frac{1}{R_{\min}^n} {}_2F_1\left(1, n; 1+n; -\frac{r_c}{R_{\min}}\right) \right] \right], \end{aligned} \quad (\text{D11})$$

where ${}_2F_1(a, b; c; x)$ is the ordinary (Gaussian) hypergeometric function.

APPENDIX E: TABLES OF USED SYMBOLS

TABLE II. Summary Greek symbols used throughout the manuscript.

Symbol	Description
$\vec{\alpha}$	Lensing deflection vector
α_{main}	Slope of the main lens galaxy projected density profile
$\vec{\alpha}_{\text{sub}}$	Deflection vector caused by substructures
$\alpha_{\text{sub},x}^{(i)}$	x-component of the net deflection caused by substructures at the i th image
β	Slope of the subhalo mass function
γ_c, γ_s	Lensing shear components
γ_{sub}	Magnitude of shear caused by substructures
δ_D^k	k -dimensional Dirac delta function
κ	Lensing convergence
κ_{sub}	Lensing convergence in mass substructures
Λ	Upper triangular matrix from the Cholesky decomposition of the covariance matrix
$\mu_s(\mathbf{x})$	Magnification tensor for the smooth lens component, evaluated at position \mathbf{x}
$\Pi(\mathbf{q})$	Prior probability distribution of \mathbf{q}
$\rho(\mathbf{r})$	Three-dimensional density profile
Σ_{crit}	Critical density for gravitational lensing
τ	$\equiv r_t/r_s$
ϕ	Projected gravitational potential
ϕ_{sub}	Substructure contribution to ϕ
$\phi_{\text{sub}}^{(i)}$	Projected gravitational potential difference between the i th image and the reference point
$\Phi_1(\mathbf{x} \mathbf{q})$	PDF of \mathbf{x} given \mathbf{q} , where \mathbf{x} is a single independent random variable
$\Phi_N(\mathbf{x} \mathbf{q})$	PDF of \mathbf{x} given \mathbf{q} , where \mathbf{x} is the sum of N independent random variables

TABLE III. Summary of roman and scripted symbols used throughout the manuscript.

Symbol	Description
a_0	Normalization of the subhalo mass function
\mathbf{A}	Transformation matrix between the linear lensing quantities and the lensing observables
\vec{A}_p, \vec{B}_p	Vectors of p th-order multipole coefficients
b	Approximate Einstein radius of main lens galaxy
\mathbf{c}_{sub}	Vector containing the individual substructure parameters
$\mathbf{c}_{\text{sub}}^{(1)}$	Substructure parameters for a single mass clump
\mathbf{C}_{sub}	Covariance matrix for the linear lensing quantities
$\mathbf{C}_{\mathbf{d}}$	Data covariance matrix
\mathbf{d}	Data vector
D_l, D_s, D_{ls}	Angular diameter distances to the lens, to the source, and from the lens to the source
h	Projection of r_{3D} along the line of sight
$\mathcal{H}_{\mathbf{d}}$	Area occupied by the distributed substructures
\mathbf{k}_L	Fourier variable conjugate to $\Delta \mathbf{t}_L$
$K_{\mathbf{p}}$	Spatial kernel for multipole expansion
$\mathcal{K}[n]$	n th-order spatial kernel for the multipole expansion of the substructure covariance matrix
$\mathcal{L}(\mathbf{x} \mathbf{d})$	Likelihood of theory vector \mathbf{x} given data vector \mathbf{d}
m	Normalized substructure mass $\equiv M_{\text{sub}}/(\pi \Sigma_{\text{crit}})$
M_0	Reference mass for subhalo mass function
M_{main}	Mass of main lens galaxy

TABLE IV. Summary of roman and scripted symbols used throughout the manuscript.

Symbol	Description
M_{min}	Low-mass bound of the subhalo mass function
M_{max}	High-mass bound of the subhalo mass function
M_{main}	Mass of main lens halo
M_{NFW}	Mass normalization of the NFW profile
M_{sub}	Substructure mass
$N, \langle N \rangle$	Number of mass substructure, average number of mass substructures
N_l, N_d	Number of local and distributed substructures
N_{img}	Number of lensed images
N_{mult}	Maximum number of multipole included in the expansion
$\vec{\mathcal{O}}_L$	Vector of stochastic random variables ($\equiv \Delta \mathbf{t}_L / m$)
\mathbf{p}	Multi-index (vector of multiple indices)
$P(\mathbf{x} \mathbf{q})$	PDF of vector \mathbf{x} given parameter vector \mathbf{q}
P_{sub}	PDF for the mass and position of substructures
$\mathcal{P}_M(M_{\text{sub}})$	PDF for the substructure mass
$\mathcal{P}_r(r)$	PDF for the spatial distribution of substructures
$\mathbf{q}_{\text{cos}}, \mathbf{q}_{\text{DM}}$	Cosmological parameters, Dark matter parameters
$\mathbf{q}_{\text{env}}, \mathbf{q}_{\text{los}}$	Lens environment parameters, Line-of-sight structure parameters
\mathbf{q}_{gal}	Macrolens parameters
\mathbf{q}_{sub}	Substructure population parameters
\mathbf{q}	$\equiv \{\mathbf{q}_{\text{gal}}, \mathbf{q}_{\text{env}}, \mathbf{q}_{\text{los}}\}$
$q_1(\mathbf{k} \mathbf{q})$	Characteristic function of $\Phi_1(\mathbf{x} \mathbf{q})$
$q_1^{\text{local}}, q_1^{\text{dist}}$	Local and distributed contribution to q_1
$\mathcal{Q}_N(\mathbf{k} \mathbf{q})$	Characteristic function of $\Phi_N(\mathbf{x} \mathbf{q})$
$\mathcal{Q}_N^{\text{local}}, \mathcal{Q}_N^{\text{dist}}$	Local and distributed contribution to \mathcal{Q}_N
r, θ	Two-dimensional polar coordinates
r_{3D}	Three-dimensional position of subhalo within the lens galaxy
r_c	Core radius of the substructures' spatial distribution
r_{max}	Radius where v_{max} occurs

(Table continued)

TABLE IV. (*Continued*)

Symbol	Description
r_{ref}	Reference radius where zero of projected potential is defined
r_s	Scale radius of subhalo
r_{sub}	Three-dimensional distance from center of subhalo
r_t	Tidal radius of subhalo
R_{ein}	Einstein radius of the lens
R_{min}	Minimum radius of the distributed substructure population
R_{max}	Maximum radius of the distributed substructure population
R_{vir}	Virial radius of the main lens galaxy
\mathbf{t}	Vector of theory observables
$\bar{\mathbf{t}}$	Contribution to \mathbf{t} from smooth mass component and environment
\mathbf{t}_L	Vector of linear lensing quantities
$\bar{\mathbf{t}}_L$	Contribution to \mathbf{t}_L from smooth mass component and environment
$\delta\mathbf{t}_L^{(i)}$	Contribution to \mathbf{t}_L from the i th mass substructure
$\Delta\mathbf{t}_L$	$\equiv \sum_i \delta\mathbf{t}_L^{(i)}$
$\Delta\mathbf{t}_L^{\text{local}}, \Delta\mathbf{t}_L^{\text{dist}}$	Local and distributed contributions to $\Delta\mathbf{t}_L$
$\Delta t^{(i)}$	Arrival time delay between image i and the leading image
\mathbf{u}	$\equiv \langle \Delta\mathbf{t}_L \rangle$
v_{max}	Maximum circular velocity of a dark matter halo
$\bar{\mathbf{x}}_i$	Unperturbed (from smooth model only) position of the i^{th} image
$\mathbf{x}^{(i)}$	Actual position of i th lensed image
x_{max}	$\equiv r_{\text{max}}/r_s$
z_{lens}	Redshift of the main lens

-
- [1] P. A. R. Ade, N. Aghanim, C. Armitage-Caplan, M. Arnaud, M. Ashdown, F. Atrio-Barandela, J. Aumont, C. Baccigalupi, A. J. Banday *et al.* (Planck Collaboration), *Astron. Astrophys.* **571**, A16 (2014).
- [2] A. G. Sánchez *et al.*, *Mon. Not. R. Astron. Soc.* **440**, 2692 (2014).
- [3] M. Kilbinger *et al.*, *Mon. Not. R. Astron. Soc.* **430**, 2200 (2013).
- [4] M. Davis, M. Lecar, C. Pryor, and E. Witten, *Astrophys. J.* **250**, 423 (1981).
- [5] G. R. Blumenthal, H. Pagels, and J. R. Primack, *Nature (London)* **299**, 37 (1982).
- [6] P. Bode, J. P. Ostriker, and N. Turok, *Astrophys. J.* **556**, 93 (2001).
- [7] J. J. Dalcanton and C. J. Hogan, *Astrophys. J.* **561**, 35 (2001).
- [8] A. R. Zentner and J. S. Bullock, *Astrophys. J.* **598**, 49 (2003).
- [9] D. Boyanovsky, H. J. de Vega, and N. Sanchez, *Phys. Rev. D* **78**, 063546 (2008).
- [10] D. Boyanovsky and J. Wu, *Phys. Rev. D* **83**, 043524 (2011).
- [11] R. E. Smith and K. Markovic, *Phys. Rev. D* **84**, 063507 (2011).
- [12] C. Boehm and R. Schaeffer, *Astron. Astrophys.* **438**, 419 (2005).
- [13] F.-Y. Cyr-Racine and K. Sigurdson, *Phys. Rev. D* **87**, 103515 (2013).
- [14] F.-Y. Cyr-Racine, R. de Putter, A. Raccanelli, and K. Sigurdson, *Phys. Rev. D* **89**, 063517 (2014).
- [15] C. Boehm, J. A. Schewtschenko, R. J. Wilkinson, C. M. Baugh, and S. Pascoli, *Mon. Not. R. Astron. Soc.* **445**, L31 (2014).
- [16] C. Dvorkin, K. Blum, and M. Kamionkowski, *Phys. Rev. D* **89**, 023519 (2014).
- [17] M. R. Buckley, J. Zavala, F.-Y. Cyr-Racine, K. Sigurdson, and M. Vogelsberger, *Phys. Rev. D* **90**, 043524 (2014).
- [18] A. Klypin, A. V. Kravtsov, O. Valenzuela, and F. Prada, *Astrophys. J.* **522**, 82 (1999).
- [19] B. Moore, S. Ghigna, F. Governato, G. Lake, T. Quinn, J. Stadel, and P. Tozzi, *Astrophys. J.* **524**, L19 (1999).
- [20] A. M. Nierenberg, T. Treu, S. A. Wright, C. D. Fassnacht, and M. W. Auger, *Mon. Not. R. Astron. Soc.* **442**, 2434 (2014).
- [21] S. Vegetti, L. V. E. Koopmans, M. W. Auger, T. Treu, and A. S. Bolton, *Mon. Not. R. Astron. Soc.* **442**, 2017 (2014).
- [22] A. M. Nierenberg, D. Oldenburg, and T. Treu, *Mon. Not. R. Astron. Soc.* **436**, 2120 (2013).
- [23] D. Xu, D. Sluse, L. Gao, J. Wang, C. Frenk, S. Mao, P. Schneider, and V. Springel, *Mon. Not. R. Astron. Soc.* **447**, 3189 (2015).

- [24] Y. Hezaveh, N. Dalal, G. Holder, M. Kuhlen, D. Marrone, N. Murray, and J. Vieira, *Astrophys. J.* **767**, 9 (2013).
- [25] C. L. MacLeod, R. Jones, E. Agol, and C. S. Kochanek, *Astrophys. J.* **773**, 35 (2013).
- [26] R. Li, H. J. Mo, Z. Fan, X. Yang, and F. C. v. d. Bosch, *Mon. Not. R. Astron. Soc.* **430**, 3359 (2013).
- [27] E. Zackrisson, S. Asadi, K. Wiik, J. Jönsson, P. Scott, K. K. Datta, M. M. Friedrich, H. Jensen, J. Johansson, C.-E. Rydberg, and A. Sandberg, *Mon. Not. R. Astron. Soc.* **431**, 2172 (2013).
- [28] S. Vegetti, D. J. Lagattuta, J. P. McKean, M. W. Auger, C. D. Fassnacht, and L. V. E. Koopmans, *Nature (London)* **481**, 341 (2012).
- [29] J. Liesenborgs and S. De Rijcke, *Mon. Not. R. Astron. Soc.* **425**, 1772 (2012).
- [30] R. Fadely and C. R. Keeton, *Mon. Not. R. Astron. Soc.* **419**, 936 (2012).
- [31] J. Chen, S. M. Koushiappas, and A. R. Zentner, *Astrophys. J.* **741**, 117 (2011).
- [32] D. D. Xu, S. Mao, A. P. Cooper, L. Gao, C. S. Frenk, R. E. Angulo, and J. Helly, *Mon. Not. R. Astron. Soc.* **421**, 2553 (2012).
- [33] A. D'Aloisio and P. Natarajan, *Mon. Not. R. Astron. Soc.* **411**, 1628 (2011).
- [34] R. B. Metcalf and A. Amara, *Mon. Not. R. Astron. Soc.* **419**, 3414 (2012).
- [35] D. D. Xu, S. Mao, A. P. Cooper, J. Wang, L. Gao, C. S. Frenk, and V. Springel, *Mon. Not. R. Astron. Soc.* **408**, 1721 (2010).
- [36] S. Vegetti and L. V. E. Koopmans, *Mon. Not. R. Astron. Soc.* **400**, 1583 (2009).
- [37] D. J. Bacon, A. Amara, and J. I. Read, *Mon. Not. R. Astron. Soc.* **409**, 389 (2010).
- [38] C. R. Keeton, arXiv:0908.3001.
- [39] D. D. Xu, S. Mao, J. Wang, V. Springel, L. Gao, S. D. M. White, C. S. Frenk, A. Jenkins, G. Li, and J. F. Navarro, *Mon. Not. R. Astron. Soc.* **398**, 1235 (2009).
- [40] A. O. Petters, B. Rider, and A. M. Teguia, *J. Math. Phys. (N.Y.)* **50**, 072503 (2009).
- [41] A. O. Petters, B. Rider, and A. M. Teguia, *J. Math. Phys. (N.Y.)* **50**, 122501 (2009).
- [42] C. R. Keeton and L. A. Moustakas, *Astrophys. J.* **699**, 1720 (2009).
- [43] S. Vegetti and L. V. E. Koopmans, *Mon. Not. R. Astron. Soc.* **392**, 945 (2009).
- [44] J. Chen, *Astron. Astrophys.* **498**, 49 (2009).
- [45] L. L. R. Williams, P. Foley, D. Farnsworth, and J. Belter, *Astrophys. J.* **685**, 725 (2008).
- [46] M. Chiba, T. Minezaki, K. T. Inoue, N. Kashikawa, H. Katata, and H. Sugai, in *Panoramic Views of Galaxy Formation and Evolution*, *Astronomical Society of the Pacific Conference Series*, Vol. 399, edited by T. Kodama, T. Yamada, and K. Aoki (Astronomical Society of the Pacific, San Francisco, 2008), p. 98.
- [47] S. Peirani, C. Alard, C. Pichon, R. Gavazzi, and D. Aubert, *Mon. Not. R. Astron. Soc.* **390**, 945 (2008).
- [48] C. Alard, *Mon. Not. R. Astron. Soc.* **388**, 375 (2008).
- [49] E. M. Shin and N. W. Evans, *Mon. Not. R. Astron. Soc.* **385**, 2107 (2008).
- [50] J. Chen, E. Rozo, N. Dalal, and J. E. Taylor, *Astrophys. J.* **659**, 52 (2007).
- [51] G. Dobler and C. R. Keeton, *Mon. Not. R. Astron. Soc.* **365**, 1243 (2006).
- [52] A. Amara, R. B. Metcalf, T. J. Cox, and J. P. Ostriker, *Mon. Not. R. Astron. Soc.* **367**, 1367 (2006).
- [53] E. Rozo, A. R. Zentner, G. Bertone, and J. Chen, *Astrophys. J.* **639**, 573 (2006).
- [54] A. V. Maccio' and M. Miranda, *Mon. Not. R. Astron. Soc.* **368**, 599 (2006).
- [55] K. T. Inoue and M. Chiba, *Astrophys. J.* **634**, 77 (2005).
- [56] M. Oguri, C. R. Keeton, and N. Dalal, *Mon. Not. R. Astron. Soc.* **364**, 1451 (2005).
- [57] N. Dalal, in *25 Years After the Discovery: Some Current Topics on Lensed QSOs*, edited by L. J. Goicoechea (2005), p. 12.
- [58] C. R. Keeton, B. S. Gaudi, and A. O. Petters, *Astrophys. J.* **635**, 35 (2005).
- [59] B. Hagan, C.-P. Ma, and A. V. Kravtsov, *Astrophys. J.* **633**, 537 (2005).
- [60] J. E. Taylor and A. Babul, *Mon. Not. R. Astron. Soc.* **364**, 535 (2005).
- [61] K. T. Inoue and M. Chiba, *Astrophys. J.* **633**, 23 (2005).
- [62] S. Mao, Y. Jing, J. Ostriker, and J. Weller, *Astrophys. J.* **604**, L5 (2004).
- [63] R. B. Metcalf, *Astrophys. J.* **622**, 72 (2005).
- [64] J. E. Taylor and A. Babul, in *Gravitational Lensing Impact on Cosmology*, *IAU Symposium*, Vol. 225, edited by Y. Mellier and G. Meylan (2005), pp. 225–230.
- [65] R. B. Metcalf, in *Gravitational Lensing Impact on Cosmology*, *IAU Symposium*, Vol. 225, edited by Y. Mellier and G. Meylan (2005), pp. 267–277.
- [66] R. Metcalf, L. Moustakas, A. Bunker, and I. Parry, *Astrophys. J.* **607**, 43 (2004).
- [67] C. Kochanek and N. Dalal, *Astrophys. J.* **610**, 69 (2004).
- [68] J. D. Cohn and C. S. Kochanek, *Astrophys. J.* **608**, 25 (2004).
- [69] M. Bradac, P. Schneider, M. Lombardi, M. Steinmetz, L. Koopmans, and J. F. Navarro, *Astron. Astrophys.* **423**, 797 (2004).
- [70] C. R. Keeton, *Astrophys. J.* **584**, 664 (2003).
- [71] C. Kochanek and N. Dalal, *AIP Conf. Proc.* **666**, 103 (2003).
- [72] L. A. Moustakas and R. B. Metcalf, *Mon. Not. R. Astron. Soc.* **339**, 607 (2003).
- [73] A. R. Zentner and J. S. Bullock, *Astrophys. J.* **598**, 49 (2003).
- [74] C. R. Keeton, B. S. Gaudi, and A. O. Petters, *Astrophys. J.* **598**, 138 (2003).
- [75] J. Chen, A. V. Kravtsov, and C. R. Keeton, *Astrophys. J.* **592**, 24 (2003).
- [76] J. E. Taylor, J. Silk, and A. Babul, in *Dark Matter in Galaxies*, *IAU Symposium*, Vol. 220, edited by S. Ryder, D. Pisano, M. Walker, and K. Freeman (2004), p. 91.
- [77] R. K. Sheth and B. Jain, *Mon. Not. R. Astron. Soc.* **345**, 529 (2003).
- [78] J. E. Taylor, A. Babul, and J. Silk, in *Satellites and Tidal Streams*, *Astronomical Society of the Pacific Conference Series*, Vol. 327, edited by F. Prada, D. Martinez Delgado,

- and T. J. Mahoney (Astronomical Society of the Pacific, San Francisco, 2004), p. 205.
- [79] N. Dalal and C. Kochanek, *Astrophys. J.* **572**, 25 (2002).
 - [80] N. Dalal and C. S. Kochanek, [arXiv:astro-ph/0202290](#).
 - [81] M. Bradač, P. Schneider, M. Steinmetz, M. Lombardi, L. J. King, and R. Porcas, *Astron. Astrophys.* **388**, 373 (2002).
 - [82] R. B. Metcalf and P. Madau, *Astrophys. J.* **563**, 9 (2001).
 - [83] C. R. Keeton, [arXiv:astro-ph/0111595](#).
 - [84] R. B. Metcalf and H. Zhao, *Astrophys. J. Lett.* **567**, L5 (2002).
 - [85] M. Chiba, *Astrophys. J.* **565**, 17 (2002).
 - [86] S. Mao and P. Schneider, *Mon. Not. R. Astron. Soc.* **295**, 587 (1998).
 - [87] L. V. E. Koopmans, *Mon. Not. R. Astron. Soc.* **363**, 1136 (2005).
 - [88] Y. Hezaveh, N. Dalal, G. Holder, T. Kisner, and M. Kuhlen, [arXiv:1403.2720](#).
 - [89] Y. D. Hezaveh, N. Dalal, D. P. Marrone, Y.-Y. Mao, W. Morningstar, D. Wen, R. D. Blandford, J. E. Carlstrom, C. D. Fassnacht, G. P. Holder, A. Kembell, P. J. Marshall, N. Murray, L. Perreault Levasseur, J. D. Vieira, and R. H. Wechsler, *Astrophys. J.* **823**, 37 (2016).
 - [90] M. Vietri and J. P. Ostriker, *Astrophys. J.* **267**, 488 (1983).
 - [91] R. Nityananda and J. P. Ostriker, *J. Astrophys. Astron.* **5**, 235 (1984).
 - [92] N. Katz, S. Balbus, and B. Paczynski, *Astrophys. J.* **306**, 2 (1986).
 - [93] P. Schneider, *Astron. Astrophys.* **179**, 80 (1987).
 - [94] P. Schneider, *Astrophys. J.* **319**, 9 (1987).
 - [95] S. Deguchi and W. D. Watson, *Phys. Rev. Lett.* **59**, 2814 (1987).
 - [96] S. Deguchi and W. D. Watson, *Astrophys. J.* **335**, 67 (1988).
 - [97] C. Seitz and P. Schneider, *Astron. Astrophys.* **288**, 1 (1994).
 - [98] C. Seitz, J. Wambsganss, and P. Schneider, *Astron. Astrophys.* **288**, 19 (1994).
 - [99] B. Neendorf, *Astron. Astrophys.* **404**, 83 (2003).
 - [100] A. V. Tuntsov and G. F. Lewis, *Mon. Not. R. Astron. Soc.* **370**, 91 (2006).
 - [101] V. V. Uchaikin and V. M. Zolotarev, *Chance and Stability: Stable Distributions and their Applications, Modern Probability and Statistics* (De Gruyter, Utrecht, 1999).
 - [102] S. Blinnikov and R. Moessner, *Astron. Astrophys. Suppl. Ser.* **130**, 193 (1998).
 - [103] P. Ade *et al.* (Planck Collaboration), [arXiv:1502.01589](#).
 - [104] J. Diemand, M. Kuhlen, and P. Madau, *Astrophys. J.* **657**, 262 (2007).
 - [105] P. Madau, J. Diemand, and M. Kuhlen, *Astrophys. J.* **679**, 1260 (2008).
 - [106] V. Springel, J. Wang, M. Vogelsberger, A. Ludlow, A. Jenkins, A. Helmi, J. F. Navarro, C. S. Frenk, and S. D. M. White, *Mon. Not. R. Astron. Soc.* **391**, 1685 (2008).
 - [107] J. Zavala, Y. P. Jing, A. Faltenbacher, G. Yepes, Y. Hoffman, S. Gottlöber, and B. Catinella, *Astrophys. J.* **700**, 1779 (2009).
 - [108] A. Zolotov, A. M. Brooks, B. Willman, F. Governato, A. Pontzen, C. Christensen, A. Dekel, T. Quinn, S. Shen, and J. Wadsley, *Astrophys. J.* **761**, 71 (2012).
 - [109] A. M. Brooks, M. Kuhlen, A. Zolotov, and D. Hooper, *Astrophys. J.* **765**, 22 (2013).
 - [110] M. Rocha, A. H. Peter, J. S. Bullock, M. Kaplinghat, S. Garrison-Kimmel, J. Onorbe, and L. A. Moustakas, *Mon. Not. R. Astron. Soc.* **430**, 81 (2013).
 - [111] A. H. G. Peter, M. Rocha, J. S. Bullock, and M. Kaplinghat, *Mon. Not. R. Astron. Soc.* **430**, 105 (2013).
 - [112] A. M. Brooks and A. Zolotov, *Astrophys. J.* **786**, 87 (2014).
 - [113] M. Vogelsberger, J. Zavala, and A. Loeb, *Mon. Not. R. Astron. Soc.* **423**, 3740 (2012).
 - [114] R. E. Angulo, O. Hahn, and T. Abel, *Mon. Not. R. Astron. Soc.* **434**, 3337 (2013).
 - [115] J. Zavala, M. Vogelsberger, and M. G. Walker, *Mon. Not. R. Astron. Soc.* **431**, L20 (2013).
 - [116] S. Garrison-Kimmel, M. Boylan-Kolchin, J. S. Bullock, and E. N. Kirby, *Mon. Not. R. Astron. Soc.* **444**, 222 (2014).
 - [117] C. Boehm, J. A. Schewtschenko, R. J. Wilkinson, C. M. Bough, and S. Pascoli, *Mon. Not. R. Astron. Soc.* **445**, L31 (2014).
 - [118] M. R. Buckley, J. Zavala, F.-Y. Cyr-Racine, K. Sigurdson, and M. Vogelsberger, *Phys. Rev. D* **90**, 043524 (2014).
 - [119] M. Vogelsberger, J. Zavala, C. Simpson, and A. Jenkins, *Mon. Not. R. Astron. Soc.* **444**, 3684 (2014).
 - [120] F. Governato, D. Weisz, A. Pontzen, S. Loebman, D. Reed, A. M. Brooks, P. Behroozi, C. Christensen, P. Madau, L. Mayer, S. Shen, M. Walker, T. Quinn, B. W. Keller, and J. Wadsley, *Mon. Not. R. Astron. Soc.* **448**, 792 (2015).
 - [121] R. T. Chamberlain, N. Dalal, A. Hearin, and P. Ricker, [arXiv:1407.2648](#).
 - [122] L. Gao, C. S. Frenk, M. Boylan-Kolchin, A. Jenkins, V. Springel, and S. D. M. White, *Mon. Not. R. Astron. Soc.* **410**, 2309 (2011).
 - [123] E. A. Baltz, P. Marshall, and M. Oguri, *J. Cosmol. Astropart. Phys.* **01** (2009) 015.
 - [124] S. Garrison-Kimmel, M. Boylan-Kolchin, J. S. Bullock, and E. N. Kirby, *Mon. Not. R. Astron. Soc.* **444**, 222 (2014).
 - [125] J. Binney and S. Tremaine, *Galactic Dynamics*, 2nd ed. (Princeton University Press, Princeton, NJ, 2008), p. 920.
 - [126] G. Tormen, A. Diaferio, and D. Syer, *Mon. Not. R. Astron. Soc.* **299**, 728 (1998).
 - [127] S. Hofmann, D. J. Schwarz, and H. Stoecker, *Phys. Rev. D* **64**, 083507 (2001).
 - [128] A. M. Green, S. Hofmann, and D. J. Schwarz, *Mon. Not. R. Astron. Soc.* **353**, L23 (2004).
 - [129] A. D. Ludlow, J. F. Navarro, V. Springel, A. Jenkins, C. S. Frenk, and Amina Helmi, *Astrophys. J.* **692**, 931 (2009).
 - [130] J. Diemand, B. Moore, and J. Stadel, *Mon. Not. R. Astron. Soc.* **352**, 535 (2004).
 - [131] T. Kollo and D. Von Rosen, *Scand. J. Stat. Theory Appl.* **25**, 93 (1998).
 - [132] C. R. Keeton, [arXiv:astro-ph/0102341](#).